

This Page Is Inserted by IFW Operations  
and is not a part of the Official Record

## **BEST AVAILABLE IMAGES**

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images may include (but are not limited to):

- BLACK BORDERS
- TEXT CUT OFF AT TOP, BOTTOM OR SIDES
- FADED TEXT
- ILLEGIBLE TEXT
- SKEWED/SLANTED IMAGES
- COLORED PHOTOS
- BLACK OR VERY BLACK AND WHITE DARK PHOTOS
- GRAY SCALE DOCUMENTS

**IMAGES ARE BEST AVAILABLE COPY.**

**As rescanning documents *will not* correct images,  
please do not report the images to the  
Image Problem Mailbox.**

**THIS PAGE BLANK (USPTO)**

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 13, 2002, 15:11:39 ; Search time 14.7 Seconds  
(without alignments)  
666.304 Million cell updates/sec

Title: US-09-631-863A-2

Perfect score: 401

Sequence: 1 MRNKRVLTKRRSGRGQD.....DNCPAKASKSPAGNSPAPL 401

Scoring table: OLIGO

Gapop 60.0 , Gapext 60.0

Searched: 231628 seqs, 2442594 residues

Word size : 0

Total number of hits satisfying chosen parameters: 231628

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : Issued Patents\_AA.\*

- 1: /cgn2\_6/ptodata/2/1aa/5A\_COMB.pep.\*
- 2: /cgn2\_6/ptodata/2/1aa/5B\_COMB.pep.\*
- 3: /cgn2\_6/ptodata/2/1aa/6A\_COMB.pep.\*
- 4: /cgn2\_6/ptodata/2/1aa/6B\_COMB.pep.\*
- 5: /cgn2\_6/ptodata/2/1aa/PCTUS\_COMB.pep.\*
- 6: /cgn2\_6/ptodata/2/1aa/backfiles1.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	9	2.2	13	4	US-09-058-562-32
2	9	2.2	15	4	US-08-602-999A-348
3	9	2.2	17	4	US-08-602-999A-413
4	9	2.2	18	4	US-08-602-999A-371
5	9	2.2	18	4	US-08-602-999A-409
6	9	2.2	19	4	US-08-602-999A-408
7	9	2.2	20	4	US-08-602-999A-357
8	9	2.2	20	4	US-08-602-999A-358
9	9	2.2	20	4	US-08-602-999A-359
10	9	2.2	20	4	US-08-602-999A-368
11	9	2.2	22	1	US-08-240-712-34
12	9	2.2	22	1	US-08-443-890-34
13	9	2.2	23	4	US-08-602-999A-376
14	9	2.2	32	3	US-08-545-196B-63
15	9	2.2	87	4	US-09-314-268-137
16	9	2.2	106	4	US-09-314-268-134
17	9	2.2	170	5	PCT-US95-16806A-6
18	9	2.2	210	1	US-08-234-783-2
19	9	2.2	210	1	US-08-456-907-2
20	9	2.2	210	5	PCT-US95-05523-2
21	9	2.2	223	1	US-08-167-035-43
22	9	2.2	223	1	US-08-208-887A-43
23	9	2.2	223	2	US-08-539-005-43
24	9	2.2	278	3	US-08-545-196B-19
25	9	2.2	281	2	US-08-810-453-2
26	9	2.2	281	3	US-08-815-190A-2
27	9	2.2	281	4	US-09-290-640-25

28	9	2.2	281	4	US-09-479-524-3	Sequence 3, Appli
29	9	2.2	281	4	US-08-339-214-8	Sequence 8, Appli
30	9	2.2	281	4	US-08-339-214-30	Sequence 30, Appli
31	9	2.2	281	5	PCT-US95-00362-2	Sequence 2, Appli
32	9	2.2	294	1	US-08-612-986-6	Sequence 6, Appli
33	9	2.2	294	1	US-08-361-806A-6	Sequence 6, Appli
34	9	2.2	294	3	US-08-545-196B-9	Sequence 4, Appli
35	9	2.2	294	4	US-09-028-327-4	Sequence 4, Appli
36	9	2.2	294	5	PCT-US95-16806A-21	Sequence 21, Appli
37	9	2.2	311	4	US-09-179-558-66	Sequence 66, Appli
38	9	2.2	377	1	US-07-863-169A-1	Sequence 1, Appli
39	9	2.2	377	2	US-08-429-964-1	Sequence 1, Appli
40	9	2.2	377	3	US-07-935-087-1	Sequence 1, Appli
41	9	2.2	377	5	PCT-US93-08062-1	Sequence 1, Appli
42	9	2.2	429	1	US-08-234-783-4	Sequence 4, Appli
43	9	2.2	429	1	US-08-456-907-4	Sequence 4, Appli
44	9	2.2	429	5	PCT-US95-05523-4	Sequence 4, Appli
45	9	2.2	445	2	US-08-900-148-2	Sequence 2, Appli

## ALIGNMENTS

RESULT 1  
US-09-058-562-32  
; Sequence 32, Application US/09058562A  
; Patent No. 6184356  
; GENERAL INFORMATION:  
; APPLICANT: Anderson, David C.  
; APPLICANT: Mathews, Antony James  
; APPLICANT: Stetler, Gary L.  
; TITLE OF INVENTION: PRODUCTION AND USE OF MULTIMERIC HEMOGLOBINS  
; FILE REFERENCE: BXTB 2087  
; CURRENT APPLICATION NUMBER: US/09/058,562A  
; CURRENT FILING DATE: 1998-04-13  
; PRIOR APPLICATION NUMBER: US 08/240,712  
; PRIOR FILING DATE: 1994-05-09  
; PRIOR APPLICATION NUMBER: PCT/US92/09752  
; PRIOR FILING DATE: 1993-05-13  
; NUMBER OF SEQ ID NOS: 33  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 32  
; LENGTH: 13  
; TYPE: PRT  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; NAME/KEY: VARIANT  
; LOCATION: (1)  
; OTHER INFORMATION: Xaa=1-3 Gly  
; NAME/KEY: VARIANT  
; LOCATION: (12)  
; OTHER INFORMATION: Xaa=1-5 Pro  
; NAME/KEY: VARIANT  
; LOCATION: (13)  
; OTHER INFORMATION: Xaa=1-3 Gly  
; OTHER INFORMATION: Description of Artificial Sequence: linker  
; NAME/KEY: HELIX  
; LOCATION: (2)-(11)  
; OTHER INFORMATION: Proline helix  
US-09-058-562-32

Query Match 2.2% Score 9; DB 4; Length 13;  
Best Local Similarity 100.0%; Pred. No. 0.071;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 48 ppppppppp 56  
|||||||  
Db 2 ppppppppp 10

RESULT 2  
US-08-602-999A-348

**THIS PAGE BLANK (USPTO)**



GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 13, 2002, 15:11:39 ; Search time 14.7 Seconds  
(without alignments)  
666.304 Million cell updates/sec

Title: US-09-631-863a-2

Perfect score: 401

Sequence: 1 MNKRVLTKRRKRRSGRGCD.....DNCPAKSKSPAGNSAPPL 401

Scoring table: OLIGO

Gapop 60.0 , Gapext 60.0

Searched: 231628 seqs, 24425594 residues

Word size : 0

Total number of hits satisfying chosen parameters: 231628

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Database :

Issued\_Patents\_AA:\*  
1: /cgn2\_6/ptodata/2/1aa/5A.COMB.pep:\*  
2: /cgn2\_6/ptodata/2/1aa/5B.COMB.pep:\*  
3: /cgn2\_6/ptodata/2/1aa/6A.COMB.pep:\*  
4: /cgn2\_6/ptodata/2/1aa/6B.COMB.pep:\*  
5: /cgn2\_6/ptodata/2/1aa/PTUS.COMB.pep:\*  
6: /cgn2\_6/ptodata/2/1aa/backfiles1.pep:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	9	2.2	13	4	US-09-058-562-32
2	9	2.2	15	4	US-08-602-999A-348
3	9	2.2	17	4	US-08-602-999A-413
4	9	2.2	18	4	US-08-602-999A-371
5	9	2.2	18	4	US-08-602-999A-409
6	9	2.2	19	4	US-08-602-999A-408
7	9	2.2	20	4	US-08-602-999A-357
8	9	2.2	20	4	US-08-602-999A-358
9	9	2.2	20	4	US-08-602-999A-359
10	9	2.2	20	4	US-08-602-999A-368
11	9	2.2	22	1	US-08-240-712-34
12	9	2.2	22	1	US-08-443-890-34
13	9	2.2	23	4	US-08-602-999A-376
14	9	2.2	32	3	US-08-545-196B-63
15	9	2.2	87	4	US-09-314-268-137
16	9	2.2	106	4	US-09-314-268-134
17	9	2.2	170	5	PCT-US95-16806A-6
18	9	2.2	210	1	US-08-234-783-2
19	9	2.2	210	1	US-08-456-907-2
20	9	2.2	210	5	PCT-US95-05523-2
21	9	2.2	223	1	US-08-167-035-43
22	9	2.2	223	1	US-08-208-887A-43
23	9	2.2	223	2	US-08-539-005-43
24	9	2.2	278	3	US-08-545-196B-19
25	9	2.2	281	3	US-08-810-453-2
26	9	2.2	281	3	US-08-815-190A-2
27	9	2.2	281	4	US-09-290-640-25

28	9	2.2	281	4	US-09-479-524-3	Sequence 3, Appli
29	9	2.2	281	4	US-08-339-214-8	Sequence 8, Appli
30	9	2.2	281	4	US-08-339-214-30	Sequence 30, Appli
31	9	2.2	281	5	PCT-US95-00362-2	Sequence 2, Appli
32	9	2.2	294	1	US-08-612-986-6	Sequence 6, Appli
33	9	2.2	294	1	US-08-361-806A-6	Sequence 6, Appli
34	9	2.2	294	4	US-08-545-196B-9	Sequence 9, Appli
35	9	2.2	294	4	US-09-028-137-4	Sequence 4, Appli
36	9	2.2	294	5	PCT-US95-16806A-21	Sequence 21, Appli
37	9	2.2	311	4	US-09-179-358-66	Sequence 66, Appli
38	9	2.2	311	1	US-07-863-169A-1	Sequence 1, Appli
39	9	2.2	377	2	US-08-429-964-1	Sequence 1, Appli
40	9	2.2	377	3	US-07-935-087-1	Sequence 1, Appli
41	9	2.2	377	5	PCT-US93-08062-1	Sequence 1, Appli
42	9	2.2	429	1	US-08-234-783-4	Sequence 1, Appli
43	9	2.2	429	1	US-08-456-907-4	Sequence 4, Appli
44	9	2.2	429	5	PCT-US95-05523-4	Sequence 4, Appli
45	9	2.2	445	2	US-08-900-148-2	Sequence 2, Appli

#### ALIGNMENTS

```
RESULT 1
US-09-058-562-32
: Sequence 32, Application US/09058562A
: Patent No. 6184356
: GENERAL INFORMATION:
: APPLICANT: Anderson, David C.
: APPLICANT: Mathews, Antony James
: APPLICANT: Stetler, Gary L.
: TITLE OF INVENTION: PRODUCTION AND USE OF MULTIMERIC HEMOGLOBINS
: FILE REFERENCE: EXTB 2087
: CURRENT APPLICATION NUMBER: US/09/058, 562A
: PRIOR FILING DATE: 1998-04-13
: PRIOR APPLICATION NUMBER: US 08/240,712
: PRIOR FILING DATE: 1994-05-09
: PRIOR APPLICATION NUMBER: PCT/US92/09752
: PRIOR FILING DATE: 1993-05-13
: NUMBER OF SEQ ID NOS: 33
: SOFTWARE: Patentin Ver. 2.1
: SEQ ID NO 32
: LENGTH: 13
: TYPE: PRT
: ORGANISM: Artificial Sequence
: FEATURE:
: NAME/KEY: VARIANT
: LOCATION: (1)
: OTHER INFORMATION: Xaa-1-3 Gly
: NAME/KEY: VARIANT
: LOCATION: (12)
: OTHER INFORMATION: Xaa-1-5 Pro
: NAME/KEY: VARIANT
: LOCATION: (13)
: OTHER INFORMATION: Xaa-1-3 Gly
: NAME/KEY: HELIX
: LOCATION: (2)..(11)
: OTHER INFORMATION: Description of Artificial Sequence: linker
: OTHER INFORMATION: Proline helix
: US-09-058-562-32

Query Match 2.28; Score 9; DB 4; Length 13;
Best Local Similarity 100.0%; Pred. No. 0.071;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 48 PPPPPPPP 56
Db 2 PPPPPPPP 10

RESULT 2
US-08-602-999A-348
```

```
; Sequence 348, Application US/08602999A
; Patent No. 6184205
; GENERAL INFORMATION:
; APPLICANT: SPARKS, Andrew B.
; APPLICANT: KAY, Brian K.
; APPLICANT: THORN, Judith M.
; APPLICANT: OQUILLIAM, Lawrence A.
; APPLICANT: DER, Channing J.
; APPLICANT: FOWLKES, Dana M.
; APPLICANT: RIDER, James E.
; TITLE OF INVENTION: SH3 BINDING PEPTIDES AND METHODS OF
; NUMBER OF SEQUENCES: 467
; TITLE OF INVENTION: ISOLATING AND USING SAME
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Pennie & Edmonds
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10036-2711
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/602,999A
; FILING DATE: 16-FEB-1996
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Mistrock, S. Leslie
; REGISTRATION NUMBER: 18,872
; REFERENCE/DOCKET NUMBER: 1101-202
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 790-9090
; TELEFAX: (212) 869-9741/8864
; TELEX: 66141 PENNIE
; INFORMATION FOR SEQ ID NO: 348:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 15 amino acids
; TYPE: amino acid
; TOPOLOGY: unknown
; MOLECULE TYPE: peptide
; US-08-602-999A-348

Query Match          2.2%: Score 9; DB 4; Length 15;
Best Local Similarity 100.0%; Pred. No. 0.081;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 48 PPPPPPPP 56
Db 1 PPPPPPPP 9
```

```
RESULT 3
US-08-602-999A-413
; Sequence 413, Application US/08602999A
; Patent No. 6184205
; GENERAL INFORMATION:
; APPLICANT: SPARKS, Andrew B.
; APPLICANT: KAY, Brian K.
; APPLICANT: THORN, Judith M.
; APPLICANT: OQUILLIAM, Lawrence A.
; APPLICANT: DER, Channing J.
; APPLICANT: FOWLKES, Dana M.
; APPLICANT: RIDER, James E.
; TITLE OF INVENTION: SH3 BINDING PEPTIDES AND METHODS OF
; NUMBER OF SEQUENCES: 467
; TITLE OF INVENTION: ISOLATING AND USING SAME
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Pennie & Edmonds
; STREET: 1155 Avenue of the Americas
```

```
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10036-2711
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/602,999A
; FILING DATE: 16-FEB-1996
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Mistrock, S. Leslie
; REGISTRATION NUMBER: 18,872
; REFERENCE/DOCKET NUMBER: 1101-202
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 790-9090
; TELEFAX: (212) 869-9741/8864
; TELEX: 66141 PENNIE
; INFORMATION FOR SEQ ID NO: 413:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 17 amino acids
; TYPE: amino acid
; TOPOLOGY: unknown
; MOLECULE TYPE: peptide
; US-08-602-999A-413

Query Match          2.2%: Score 9; DB 4; Length 17;
Best Local Similarity 100.0%; Pred. No. 0.091;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 48 PPPPPPPP 56
Db 5 PPPPPPPP 13
```

```
RESULT 4
US-08-602-999A-371
; Sequence 371, Application US/08602999A
; Patent No. 6184205
; GENERAL INFORMATION:
; APPLICANT: SPARKS, Andrew B.
; APPLICANT: KAY, Brian K.
; APPLICANT: THORN, Judith M.
; APPLICANT: OQUILLIAM, Lawrence A.
; APPLICANT: DER, Channing J.
; APPLICANT: FOWLKES, Dana M.
; APPLICANT: RIDER, James E.
; TITLE OF INVENTION: SH3 BINDING PEPTIDES AND METHODS OF
; NUMBER OF SEQUENCES: 467
; TITLE OF INVENTION: ISOLATING AND USING SAME
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Pennie & Edmonds
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10036-2711
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/602,999A
; FILING DATE: 16-FEB-1996
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Mistrock, S. Leslie
; REGISTRATION NUMBER: 18,872
```

REFERENCE/DOCKET NUMBER: 1101-202  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (212) 790-9090  
TELEFAX: (212) 869-9741/8864  
TELEX: 66141 PENNIE  
INFORMATION FOR SEQ ID NO: 371:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 18 amino acids  
TYPE: amino acid  
TOPOLOGY: unknown  
MOLECULE TYPE: peptide  
US-08-602-999A-371

Query Match 2.2%; Score 9; DB 4; Length 18;  
Best Local Similarity 100.0%; Pred. No. 0.096;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 48 PPPPPPPP 56  
|||||  
DB 5 PPPPPPPP 13

RESULT 5  
US-08-602-999A-409  
Sequence 409, Application US/08602999A  
Patent No. 6184205  
GENERAL INFORMATION:  
APPLICANT: SPARKS, Andrew B.  
APPLICANT: KAY, Brian K.  
APPLICANT: THORN, Judith M.  
APPLICANT: OUILIAM, Lawrence A.  
APPLICANT: DER, Channing J.  
APPLICANT: FOWLES, Dana M.  
APPLICANT: RIDER, James E.  
TITLE OF INVENTION: SH3 BINDING PEPTIDES AND METHODS OF  
ISOLATING AND USING SAME  
NUMBER OF SEQUENCES: 467  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Pennie & Edmonds  
STREET: 1155 Avenue of the Americas  
CITY: New York  
STATE: New York  
COUNTRY: U.S.A.  
ZIP: 10036-2711  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/602,999A  
FILING DATE: 16-FEB-1996  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Mistock, S. Leslie  
REGISTRATION NUMBER: 18,872  
REFERENCE/DOCKET NUMBER: 1101-202  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (212) 790-9090  
TELEFAX: (212) 869-9741/8864  
TELEX: 66141 PENNIE  
INFORMATION FOR SEQ ID NO: 409:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 18 amino acids  
TYPE: amino acid  
TOPOLOGY: unknown  
MOLECULE TYPE: peptide  
US-08-602-999A-409

Query Match 2.2%; Score 9; DB 4; Length 18;  
Best Local Similarity 100.0%; Pred. No. 0.096;

Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 48 PPPPPPPP 56  
|||||  
DB 5 PPPPPPPP 13

RESULT 6  
US-08-602-999A-408  
Sequence 408, Application US/08602999A  
Patent No. 6184205  
GENERAL INFORMATION:  
APPLICANT: SPARKS, Andrew B.  
APPLICANT: KAY, Brian K.  
APPLICANT: THORN, Judith M.  
APPLICANT: OUILIAM, Lawrence A.  
APPLICANT: DER, Channing J.  
APPLICANT: FOWLES, Dana M.  
APPLICANT: RIDER, James E.  
TITLE OF INVENTION: SH3 BINDING PEPTIDES AND METHODS OF  
ISOLATING AND USING SAME  
NUMBER OF SEQUENCES: 467  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Pennie & Edmonds  
STREET: 1155 Avenue of the Americas  
CITY: New York  
STATE: New York  
COUNTRY: U.S.A.  
ZIP: 10036-2711  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/602,999A  
FILING DATE: 16-FEB-1996  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Mistock, S. Leslie  
REGISTRATION NUMBER: 18,872  
REFERENCE/DOCKET NUMBER: 1101-202  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (212) 790-9090  
TELEFAX: (212) 869-9741/8864  
TELEX: 66141 PENNIE  
INFORMATION FOR SEQ ID NO: 408:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 19 amino acids  
TYPE: amino acid  
TOPOLOGY: unknown  
MOLECULE TYPE: peptide  
US-08-602-999A-408

Query Match 2.2%; Score 9; DB 4; Length 19;  
Best Local Similarity 100.0%; Pred. No. 0.1;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 48 PPPPPPPP 56  
|||||  
DB 5 PPPPPPPP 13

RESULT 7  
US-08-602-999A-357  
Sequence 357, Application US/08602999A  
Patent No. 6184205  
GENERAL INFORMATION:  
APPLICANT: SPARKS, Andrew B.  
APPLICANT: KAY, Brian K.  
APPLICANT: THORN, Judith M.  
APPLICANT: OUILIAM, Lawrence A.

APPLICANT: DER, Channing J.  
APPLICANT: FOWLES, Dana M.  
APPLICANT: RIDER, James E.  
TITLE OF INVENTION: SH3 BINDING PEPTIDES AND METHODS OF  
TITLE OF INVENTION: ISOLATING AND USING SAME  
NUMBER OF SEQUENCES: 467  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Pennie & Edmonds  
STREET: 1155 Avenue of the Americas  
CITY: New York  
STATE: New York  
COUNTRY: U.S.A.  
ZIP: 10036-2711  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/602,999A  
FILING DATE: 16-FEB-1996  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Mistrock, S. Leslie  
REGISTRATION NUMBER: 18,872  
REFERENCE/DOCKET NUMBER: 1101-202  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (212) 790-9090  
TELEFAX: (212) 869-9741/8864  
TELEX: 66141 PENNIE  
INFORMATION FOR SEQ ID NO: 357:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 20 amino acids  
TYPE: amino acid  
TOPOLOGY: unknown  
MOLECULE TYPE: peptide  
US-08-602-999A-357

Query Match 2.2%; Score 9; DB 4; Length 20;  
Best Local Similarity 100.0%; Pred. No. 0.11;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 48 PPPPPPPP 56  
|||||  
DB 5 PPPPPPPP 13

RESULT 8  
US-08-602-999A-358  
Sequence 358, Application US/08602999A  
Patent No. 6184205  
GENERAL INFORMATION:  
APPLICANT: SPARKS, Andrew B.  
APPLICANT: KAY, Brian K.  
APPLICANT: THORN, Judith M.  
APPLICANT: QUILIAM, Lawrence A.  
APPLICANT: DER, Channing J.  
APPLICANT: FOWLES, Dana M.  
APPLICANT: RIDER, James E.  
TITLE OF INVENTION: SH3 BINDING PEPTIDES AND METHODS OF  
TITLE OF INVENTION: ISOLATING AND USING SAME  
NUMBER OF SEQUENCES: 467  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Pennie & Edmonds  
STREET: 1155 Avenue of the Americas  
CITY: New York  
STATE: New York  
COUNTRY: U.S.A.  
ZIP: 10036-2711  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/602,999A  
FILING DATE: 16-FEB-1996  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Mistrock, S. Leslie  
REGISTRATION NUMBER: 18,872  
REFERENCE/DOCKET NUMBER: 1101-202  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (212) 790-9090  
TELEFAX: (212) 869-9741/8864  
TELEX: 66141 PENNIE  
INFORMATION FOR SEQ ID NO: 358:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 20 amino acids  
TYPE: amino acid  
TOPOLOGY: unknown  
MOLECULE TYPE: peptide  
US-08-602-999A-358

Query Match 2.2%; Score 9; DB 4; Length 20;  
Best Local Similarity 100.0%; Pred. No. 0.11;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 48 PPPPPPPP 56  
|||||  
DB 5 PPPPPPPP 13

RESULT 9  
US-08-602-999A-359  
Sequence 359, Application US/08602999A  
Patent No. 6184205  
GENERAL INFORMATION:  
APPLICANT: SPARKS, Andrew B.  
APPLICANT: KAY, Brian K.  
APPLICANT: THORN, Judith M.  
APPLICANT: QUILIAM, Lawrence A.  
APPLICANT: DER, Channing J.  
APPLICANT: FOWLES, Dana M.  
APPLICANT: RIDER, James E.  
TITLE OF INVENTION: SH3 BINDING PEPTIDES AND METHODS OF  
TITLE OF INVENTION: ISOLATING AND USING SAME  
NUMBER OF SEQUENCES: 467  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Pennie & Edmonds  
STREET: 1155 Avenue of the Americas  
CITY: New York  
STATE: New York  
COUNTRY: U.S.A.  
ZIP: 10036-2711  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/602,999A  
FILING DATE: 16-FEB-1996  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Mistrock, S. Leslie  
REGISTRATION NUMBER: 18,872  
REFERENCE/DOCKET NUMBER: 1101-202  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (212) 790-9090  
TELEFAX: (212) 869-9741/8864  
TELEX: 66141 PENNIE  
INFORMATION FOR SEQ ID NO: 359:  
SEQUENCE CHARACTERISTICS:

LENGTH: 20 amino acids  
TYPE: amino acid  
TOPOLOGY: unknown  
MOLECULE TYPE: peptide  
US-08-602-999A-359

Query Match 2.2%; Score 9; DB 4; Length 20;  
Best Local Similarity 100.0%; Pred. No. 0.11;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 48 PPPPPPPP 56  
|||||  
DB 5 PPPPPPPP 13

RESULT 10  
US-08-602-999A-368  
Sequence 368, Application US/08602999A  
Patent No. 6184205  
GENERAL INFORMATION:  
APPLICANT: SPARKS, Andrew B.  
APPLICANT: KAY, Brian K.  
APPLICANT: THORN, Judith M.  
APPLICANT: OUILLIAM, Lawrence A.  
APPLICANT: DER, Channing J.  
APPLICANT: FOWLES, Dana M.  
APPLICANT: RIDER, James E.  
TITLE OF INVENTION: SH3 BINDING PEPTIDES AND METHODS OF  
TITLE OF INVENTION: ISOLATING AND USING SAME  
NUMBER OF SEQUENCES: 467  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Pennie & Edmonds  
STREET: 1155 Avenue of the Americas  
CITY: New York  
STATE: New York  
COUNTRY: U.S.A.  
ZIP: 10036-2711  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/602,999A  
FILING DATE: 16-FEB-1996  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: MISTOCK, S. Leslie  
REGISTRATION NUMBER: 18,872  
REFERENCE/DOCKET NUMBER: 1101-202  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (212) 790-9090  
TELEFAX: (212) 869-9741/8864  
TELEX: 66141 PENNIE  
INFORMATION FOR SEQ ID NO: 368:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 20 amino acids  
TYPE: amino acid  
TOPOLOGY: unknown  
MOLECULE TYPE: peptide  
US-08-602-999A-368

Query Match 2.2%; Score 9; DB 4; Length 20;  
Best Local Similarity 100.0%; Pred. No. 0.11;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 48 PPPPPPPP 56  
|||||  
DB 4 PPPPPPPP 12

RESULT 11  
US-08-240-712-34  
Sequence 34, Application US/08240712  
Patent No. 5599607  
GENERAL INFORMATION:  
APPLICANT: ANDERSON, DAVID C.  
APPLICANT: MATHEWS, ANTONY JAMES  
APPLICANT: STETLER, GARY L.  
TITLE OF INVENTION: PRODUCTION AND USE OF MULTIMERIC  
TITLE OF INVENTION: HEMOGLOBINS  
NUMBER OF SEQUENCES: 35  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Browdy and Neimark  
STREET: 419 Seventh Street, N.W., Suite 300  
CITY: Washington  
STATE: D.C.  
COUNTRY: USA  
ZIP: 20004  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/240,712  
FILING DATE: 09-MAY-1994  
CLASSIFICATION: 530  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: PCT/US92/09752  
FILING DATE: 13-MAY-1993  
ATTORNEY/AGENT INFORMATION:  
NAME: COOPER, IVER P  
REGISTRATION NUMBER: 28,005  
REFERENCE/DOCKET NUMBER: ANDERSON-6  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 202-628-5197  
TELEFAX: 202-737-3528  
TELEX: 248633  
INFORMATION FOR SEQ ID NO: 34:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 22 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
FEATURE:  
OTHER INFORMATION: /note= one or both of Gly  
OTHER INFORMATION: residues 2 and 3 can be absent; any or all  
OTHER INFORMATION: of Pro residues 16, 17, 18 and 19 can be  
OTHER INFORMATION: absent; one or both of Gly residues 22 and  
OTHER INFORMATION: 23 can be absent  
US-08-240-712-34

Query Match 2.2%; Score 9; DB 1; Length 22;  
Best Local Similarity 100.0%; Pred. No. 0.12;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 48 PPPPPPPP 56  
|||||  
DB 4 PPPPPPPP 12

RESULT 12  
US-08-443-890-34  
Sequence 34, Application US/08443890  
Patent No. 5739011  
GENERAL INFORMATION:  
APPLICANT: ANDERSON, DAVID C.  
APPLICANT: MATHEWS, ANTONY JAMES  
APPLICANT: STETLER, GARY L.  
TITLE OF INVENTION: PRODUCTION AND USE OF MULTIMERIC  
TITLE OF INVENTION: HEMOGLOBINS

NUMBER OF SEQUENCES: 35  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Browdy and Neimark  
STREET: 419 Seventh Street, N.W., Suite 300  
CITY: Washington  
STATE: D.C.  
COUNTRY: USA  
ZIP: 20004  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentln Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/443,890  
FILING DATE: 31-MAY-1995  
CLASSIFICATION: 530  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/240,712  
FILING DATE: 09-MAY-1994  
APPLICATION NUMBER: PCT/US92/09752  
FILING DATE: 13-MAY-1993  
ATTORNEY/AGENT INFORMATION:  
NAME: COOPER, TYER P  
REGISTRATION NUMBER: 28,005  
REFERENCE/DOCKET NUMBER: ANDERSON-6  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 202-628-5197  
TELEFAX: 202-737-3528  
TELEX: 248633  
INFORMATION FOR SEQ ID NO: 34:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 22 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
FEATURE:  
OTHER INFORMATION: /note= one or both of Gly  
OTHER INFORMATION: residues 2 and 3 can be absent; any or all  
OTHER INFORMATION: of Pro residues 16, 17, 18 and 19 can be  
OTHER INFORMATION: absent; one or both of Gly residues 22 and  
OTHER INFORMATION: 23 can be absent  
US-08-443-890-34  
Query Match 2.2%; Score 9; DB 1; Length 22;  
Best Local Similarity 100.0%; Pred. No. 0.12;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 48 PPPPPPPP 56  
|||||  
DB 4 PPPPPPPP 12

RESULT 13  
US-08-602-999A-376  
Sequence 376, Application US/08602999A  
Patent No. 6184205  
GENERAL INFORMATION:  
APPLICANT: SPARKS, Andrew B.  
APPLICANT: KAY, Brian K.  
APPLICANT: THORN, Judith M.  
APPLICANT: OULLIAM, Lawrence A.  
APPLICANT: DER, Channing J.  
APPLICANT: FOWKES, Dana M.  
APPLICANT: RIDER, James E.  
TITLE OF INVENTION: SH3 BINDING PEPTIDES AND METHODS OF  
TITLE OF INVENTION: ISOLATING AND USING SAME  
NUMBER OF SEQUENCES: 467  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Pennie & Edmonds  
STREET: 1155 Avenue of the Americas

CITY: New York  
STATE: New York  
COUNTRY: U.S.A.  
ZIP: 10036-2711  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentln Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/602,999A  
FILING DATE: 16-FEB-1996  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Mistrock, S. Leslie  
REGISTRATION NUMBER: 18,872  
REFERENCE/DOCKET NUMBER: 1101-202  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (212) 790-9090  
TELEFAX: (212) 869-9741/8864  
TELEX: 66141 PENNIE  
INFORMATION FOR SEQ ID NO: 376:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 23 amino acids  
TYPE: amino acid  
TOPOLOGY: unknown  
MOLECULE TYPE: peptide  
US-08-602-999A-376

Query Match 2.2%; Score 9; DB 4; Length 23;  
Best Local Similarity 100.0%; Pred. No. 0.12;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 48 PPPPPPPP 56  
|||||  
DB 5 PPPPPPPP 13

RESULT 14  
US-08-545-196B-63  
Sequence 63, Application US/08545196B  
Patent No. 6080577  
GENERAL INFORMATION:  
APPLICANT: MELKI, JUDITH  
APPLICANT: MURNICH, ARNOLD  
TITLE OF INVENTION: SURVIVAL MOTOR NEURON (SMN) GENE: A GENE  
TITLE OF INVENTION: FOR SPINAL MUSCULAR ATROPHY  
NUMBER OF SEQUENCES: 65  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: BIRCH, STEWART, KOLASCH AND BIRCH, LLP  
STREET: PO BOX 747  
CITY: FALLS CHURCH  
STATE: VA  
COUNTRY: USA  
ZIP: 22040-0747  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentln Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/545,196B  
FILING DATE: 19-OCT-1995  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: FARACI, C. J.  
REGISTRATION NUMBER: 32,350  
REFERENCE/DOCKET NUMBER: 2121-110P  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (703) 205-8000  
TELEFAX: (703) 205-8050  
INFORMATION FOR SEQ ID NO: 63:

SEQUENCE CHARACTERISTICS:  
LENGTH: 32 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-545-196B-63

Query Match 2.2%; Score 9; DB 3; Length 32;  
Best Local Similarity 100.0%; Pred. No. 0.16;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 48 PPPPPPPP 56  
|||||  
DB 8 PPPPPPPP 16

RESULT 15  
US-09-314-268-137  
; Sequence 137, Application US/09314268  
; Patent No. 6346377  
; GENERAL INFORMATION:  
; APPLICANT: DOODAR, John  
; TITLE OF INVENTION: IMPROVEMENTS IN OR RELATING TO SCREENING FOR PAPILLOMA  
; FILE REFERENCE: 3789/80902  
; CURRENT APPLICATION NUMBER: US/09/314,268  
; CURRENT FILING DATE: 1999-03-19  
; EARLIER APPLICATION NUMBER: 09/314,268  
; EARLIER FILING DATE: 1999-05-18  
; NUMBER OF SEQ ID NOS: 179  
; SOFTWARE: Patentln Ver. 2.1  
; SEQ ID NO 137  
; LENGTH: 87  
; TYPE: PRT  
; ORGANISM: Human papillomavirus type 36  
US-09-314-268-137

Query Match 2.2%; Score 9; DB 4; Length 87;  
Best Local Similarity 100.0%; Pred. No. 0.41;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 48 PPPPPPPP 56  
|||||  
DB 77 PPPPPPPP 85

Search completed: August 13, 2002, 15:13:57  
Job time: 138 sec

**THIS PAGE BLANK (USPTO)**



GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 13, 2002, 15:07:09 ; Search time 31.61 Seconds  
(without alignments)  
1409.067 Million cell updates/sec

Title: US-09-631-863A-2  
Perfect score: 2139  
Sequence: 1 MRNKRVLTKTKRRSGRGGD.....DNCPRKSSPAGNSPAPL 401

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 747574 seqs, 111073796 residues

Total number of hits satisfying chosen parameters: 747574

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

- A\_Geneseq\_032802.\*
- 1: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1980.DAT.\*
  - 2: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1981.DAT.\*
  - 3: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1982.DAT.\*
  - 4: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1983.DAT.\*
  - 5: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1984.DAT.\*
  - 6: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1985.DAT.\*
  - 7: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1986.DAT.\*
  - 8: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1987.DAT.\*
  - 9: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1988.DAT.\*
  - 10: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1989.DAT.\*
  - 11: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1990.DAT.\*
  - 12: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1991.DAT.\*
  - 13: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1992.DAT.\*
  - 14: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1993.DAT.\*
  - 15: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1994.DAT.\*
  - 16: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1995.DAT.\*
  - 17: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1996.DAT.\*
  - 18: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1997.DAT.\*
  - 19: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1998.DAT.\*
  - 20: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1999.DAT.\*
  - 21: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA2000.DAT.\*
  - 22: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA2001.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1802	84.2	342	21	Human ORFX ORF1912
2	245.5	11.5	1607	22	Novel human diagno
3	191	8.9	502	22	Novel human diagno
4	159	7.4	146	20	Human alpha helica
5	159	7.4	146	22	Human cell cycle a
6	156	7.3	520	22	Protein encoded by
7	155	7.2	126	20	Human mature alpha
8	149.5	7.0	110	20	Human zalphal epit
9	149.5	7.0	168	21	Human zsig83 matur
10	149.5	7.0	184	21	Human zsig83 prote
11	139.5	6.5	142	22	Novel human diagno

12	139	6.5	113	22	AA60475	Human cell cycle a
13	137	6.4	144	20	AA55927	Human myometrium t
14	135.5	6.3	1230	19	AAW48895	Candida albicans C
15	132	6.2	557	19	AAW20666	Human neurofilamen
16	130	6.1	676	22	AB59726	Drosophila melanog
17	129	6.0	872	22	AB90753	Human shear stress
18	126.5	5.9	567	22	AB64078	Drosophila melanog
19	126.5	5.9	567	22	AB65901	Drosophila melanog
20	126.5	5.9	567	22	AB66007	Drosophila melanog
21	126	5.9	1072	22	AA670871	C albicans apoptos
22	125	5.8	905	18	AAW31186	Human p160 polyep
23	125	5.8	1135	18	AAW31185	Human p160 polyep
24	123	5.8	874	22	AB62601	Drosophila melanog
25	123	5.8	1321	22	AB625416	Novel human diagno
26	122.5	5.7	1316	22	AB622997	Novel human diagno
27	122.5	5.7	2348	22	AB610929	Novel human diagno
28	121.5	5.7	486	22	AAU30801	Novel human secret
29	121.5	5.7	529	22	AB612368	Human bone marrow
30	120.5	5.6	1157	22	AB619976	Novel human diagno
31	120.5	5.6	1743	22	AB610928	Novel human diagno
32	120	5.6	80	20	AA626056	Human zalphal epit
33	120	5.6	954	22	AAU14615	Novel bone marrow
34	119.5	5.6	622	22	AB62816	Drosophila melanog
35	119.5	5.6	2954	20	AAU01632	Amino acid sequenc
36	119	5.6	505	19	AAW46889	Human Neural-Wisko
37	119	5.6	505	22	AA652316	Human N-WASP prote
38	119	5.6	505	22	AA657338	Amino acid sequenc
39	119	5.6	582	22	AB616505	Novel human diagno
40	118.5	5.5	2293	21	AA659197	Amino acid sequenc
41	118	5.5	818	18	AAW13386	Human protein ubiq
42	118	5.5	2175	22	AB665698	Drosophila melanog
43	117.5	5.5	501	19	AAW46890	Rat Neural-Wiskott
44	117.5	5.5	501	22	AA652319	Rat N-WASP protein
45	117.5	5.5	501	22	AA667355	Amino acid sequenc

ALIGNMENTS

RESULT	1
AA642148	
ID	AA642148 standard; Protein: 342 AA.
XX	
AC	AA642148;
XX	
DT	08-FEB-2001 (first entry)
XX	
DE	Human ORFX ORF1912 polypeptide sequence SEQ ID NO:3824.
XX	
KW	Human: open reading frame; ORFX: detection; cytotropic; hepatotropic;
KW	vulnerable; antipariatic; antipariatic; antipariatic; antipariatic;
KW	anticonvulsant; osteopathic; antidiabetic; immunosuppressant; cardiac;
KW	immunostimulant; thrombolytic; coagulant; vasotropic; antidiabetic;
KW	hypotensive; dermatological; immunosuppressive; antitumor;
KW	antiviral; antibacterial; antifungal; antitumor; antitumor;
KW	antitumor; gene therapy; cancer; proliferative disorder; hypertension;
KW	neurodegenerative disorder; osteoarthritis; graft vs host disease;
KW	cardiovascular disease; diabetes mellitus; hypothyroidism; SCID; AIDS;
KW	cholesterol ester storage; systemic lupus erythematosus; infection;
KW	severe combined immunodeficiency; malaria; autoimmune disorder; asthma;
KW	allergy; aplastic anaemia; nocturnal haemoglobinuria; burn; wound;
KW	bone damage; cartilage damage; antiinflammatory disease; coagulation;
KW	thrombosis; contraceptive.
XX	
OS	Homo sapiens.
XX	
PN	WO200058473-A2.
XX	
PD	05-OCT-2000.
XX	
PF	31-MAR-2000; 2000WO-US08621.
XX	
PR	31-MAR-1999; 99US-0127607.



QY 100 SEENNLSQ-----QVQLTEENTT-----LREQVEPTPEDEDD 134  
Db 385 qeatldlkgreetpqeqnqtehtaelmamvrtsiislyfrmqdlkeq-----grvae 439  
QY 135 IELGAAAAAPPPIEECEPEDLPKFDGNDPMDLAFPMACQIFMEKSTRDFSVDVRV 194  
Db 440 ilkginagqlp-----apkhfgsgdrrefhefivlcqitlqsgyrfmfindrlv 488  
QY 195 CFVTSMTGTGAARWASAKLERSHYLMHNYPAFMEMKHVFEDPQRRVAKRKIRRLRQGM 254  
Db 489 gyvinhlsglalewakallgenspligdfpafleamsevfeyrqalrvaeamftrqgg 548  
QY 255 GSVIDYNAFQMIQADLDWNEPALIDQYHEGLSDHIQIELSHLEVAKSLSALIGQCIIH-I 313  
Db 549 rsateyidefgslvpilgwpdevlqahlcqglneerhyl--frvpqdsrrqsdcapha 606  
QY 314 ERRLLA---RAAAARKPRSPRALVPHIASHHQVD 345  
Db 607 nrreagreksyaqaagrppepdldrltcsrevd 641  
RESULT 3  
ABG09602  
ID ABG09602 standard; Protein; 502 AA.  
XX AC  
XX ABG09602;  
DT 13-FEB-2002 (first entry)  
XX DE  
XX Novel human diagnostic protein #9593.  
XX KW Human; chromosome mapping; gene mapping; gene therapy; forensic;  
XX KW food supplement; medical imaging; diagnostic; genetic disorder.  
XX OS Homo sapiens.  
XX PN WO200175067-A2.  
XX PD 11-OCT-2001.  
XX PF 30-MAR-2001; 2001WO-US08631.  
XX PR 31-MAR-2000; 2000US-0540217.  
XX PR 23-AUG-2000; 2000US-0649167.  
XX PA (HYSE-) HYSEQ INC.  
XX PI Drmanac RT, Liu C, Tang YT;  
XX DR WPI; 2001-639362/73.  
XX DR N-PSDB; AAS73789.  
XX PT New isolated polynucleotide and encoded polypeptides, useful in  
PT diagnostics, forensics, gene mapping, identification of mutations  
PT responsible for genetic disorders or other traits and to assess  
PT biodiversity -  
XX PS Claim 20; SEQ ID NO 39961; 103pp; English.  
XX CC The invention relates to isolated polynucleotide (I) and  
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,  
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome  
CC and gene mapping, and in recombinant production of (II). The  
CC polynucleotides are also used in diagnostics as expressed sequence tags  
CC for identifying expressed genes. (I) is useful in gene therapy techniques  
CC to restore normal activity of (II) or to treat disease states involving  
CC (II). (II) is useful for generating antibodies against it, detecting or  
CC quantitating a polypeptide in tissue, as molecular weight markers and as  
CC a food supplement. (II) and its binding partners are useful in medical  
CC imaging of sites expressing (II). (I) and (II) are useful for treating  
CC disorders involving aberrant protein expression or biological activity.  
CC The polypeptide and polynucleotide sequences have applications in  
CC diagnostics, forensics, gene mapping, identification of mutations

CC responsible for genetic disorders or other traits to assess biodiversity  
CC and to produce other types of data and products dependent on DNA and  
CC amino acid sequences. ABG00010-ABG30377 represent novel human  
CC diagnostic amino acid sequences of the invention.  
CC Note: The sequence data for this patent did not appear in the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pct\_sequences.  
XX SQ Sequence 502 AA;  
Query Match 8.9%; Score 191; DB 22; Length 502;  
Best Local Similarity 25.2%; Pred. No. 4.4e-07;  
Matches 55; Conservative 43; Mismatches 110; Indels 10; Gaps 3;  
QY 158 LPEKFDGNDPMDLAFPMACQIFMEKSTRDFSVDVRVCFVTSMTGTGAARWASAKLERSH 217  
Db 21 lplragspglmpveeemeideekgmkgflddsermaflvslhlgaerwflilqmevge 80  
QY 218 YLMHNYPAFMEMKHVFEDPQRRVAKRKIRRLRQGMGVSVIDYNAFQMIQADLDWNEPA 277  
Db 81 plshenksfrrsggiydsidlsavichpkqgksvryatdfillarhiswsdai 140  
QY 278 LIDQYHEGLSDHIQIELSH--LEVAKSLSALIGQCIIHIERLARAAAAARKPRSPRALVL 335  
Db 141 lrtrfleglseavtkmgrriflkvagslkelidrslytecqlae----ekdspgnssqvl 196  
QY 336 PHIASHHQVDTEPVGGARMRLTQEEKERRRKLNLCLY 373  
Db 197 ptaekrn----neeamgneissqgqteehqhvkrcyy 230  
RESULT 4  
AAY26051  
ID AAY26051 standard; Protein; 146 AA.  
XX AC  
XX AAY26051;  
DT 28-SEP-1999 (first entry)  
XX DE Human alpha helical precursor protein-1, zalphal.  
XX KW Human alpha helical precursor protein-1; zalphal; helical cytokine;  
KW growth hormone; erythropoietin; leptin; chromosome Xq27.3; FMRI;  
KW Fragile-X syndrome; interleukin-10; connective tissue dysfunction;  
KW abnormal proliferative disorder; cancer; epidermal system; skin tone;  
KW hypothalamic pituitary gonadal axis dysfunction; cardiovascular system;  
KW cosmetic improvement; elasticity.  
XX OS Homo sapiens.  
XX FH Key  
XX Peptide Location/Qualifiers  
FT 1..20 /label= signal\_peptide  
FT 21..146 /label= zalphal  
FT 23..37 /label= Helix\_A  
FT 53..67 /label= Helix\_B  
FT 82..96 /label= Helix\_C  
FT 118..132 /label= Helix\_D  
XX WO9929720-A2.  
XX PN 17-JUN-1999.  
XX PD 10-DEC-1998; 98WO-US26273.  
XX PR 10-DEC-1997; 97US-0987926.



KW Gene identification; essential gene; GRACE; pathogenic fungus;  
KW gene replacement and conditional expression; fungal infection.

OS Candida albicans.

PN WO200160975-A2.

PD 23-AUG-2001.

PF 20-FEB-2001; 2001WO-US05551.

PR 18-FEB-2000; 2000US-0183534.

XX (ELIT-) ELITRA PHARM INC.

PI Roemer T, Jiang B, Boone C, Bussey H;

XX WPI; 2001-489080/53.

DR N-PSDB; AAS23439.

XX Identifying genes essential to fungal metabolisms and identifying  
PT potential therapeutic agents that target these genes -

XX Claim 43; Page 247-248; 324pp; English.

CC The present invention relates to novel methods for constructing fungal  
CC strains useful for identification and validation of gene products as  
CC targets for therapeutic agents, for creating a collection of identified  
CC essential genes, and screening assays for the discovery of new drugs.  
CC The invention provides the GRACE (gene replacement and conditional  
CC expression) method for the construction of mutant organisms referred to  
CC as GRACE strains of the organism. The invention can be applied to any  
CC organism, particularly a pathogenic fungus e.g. Candida albicans,  
CC Aspergillus fumigatus and Cryptococcus neoformans. The methods are  
CC useful to identify agents that may be used in the treatment of fungal  
CC infections. AAU15053-AAU15113 represent proteins encoded by C. albicans  
CC essential genes.

XX Sequence 520 AA;

Query Match 7.3%; Score 156; DB 22; Length 520;

Best Local Similarity 21.6%; Pred. No. 0.00031;

Matches 78; Conservative 59; Mismatches 118; Indels 106; Gaps 18;

QY 9 TKRRSGRGQDGLPHRS-----EATAGRSPPR--PTVTILGPCPPP-----50

Db 24 tkllksptayhkpshvherhsitsmldtpsdtpktkptptisfrkpsislltspsva 83

QY 51 --PPPPPPNNNNNNKHTGHSACVNNNTERR-----RDELSEETNNLRE 94

Db 84 hkppplpsssvsgssehsarss--paiknrsianiidayeeepatktekkaelnspk- 140

QY 95 KVMKQS-----EENNLSQVOKLTENTTLREQVEPTP--EDEDDDI-ELRGAAAAA 145

Db 141 --indstpvpkLeehendtnkvekvdvsapepkpkkeqpvdqddldtklkkqskk 198

QY 146 P----PPPI-----EECEDPELPEKFDGPDMLAPMAOQIPMEKSTDFSV 190

Db 199 prryetpplwaqrvwpnrqrkeetnvd-----dgneaitr--lsekpvfdytrtsvdl 251

QY 191 RVRVCFTVSM-----MTGRAAWASAKLERSHYLMHNPAPFMWEMKHFDPQRREVA 245

Db 252 ----csitgmippssitrkiaewyvanfsn-----veekskrnv--- 286

QY 246 KTRLRQMGSGVIDYSNAPOMTAODLWNEPALIDQYHEGLSDHIQEELSHLEVAKLSA 305

Db 287 ----eclkfgkiidkrsgnri---dlnvvtectiftdhssvffdmqveevawkeikfkld 340

QY 306 L 306

Db 341 L 341

RESULT 7

AAV25512

ID AAV25512 standard; Protein; 126 AA.

XX AC AAV25512;

XX 28-SEP-1999 (first entry)

XX Human mature alpha helical protein-1 zalphal.

XX Human alpha helical protein-1; zalphal; helical cytokine; skin tone;  
KW growth hormone; erythropoietin; leptin; interleukin-10; gene therapy;  
KW chromosome Xq27.3; FMR1; Fragile-X syndrome; cosmetic improvement;  
KW hypothalamic-pituitary gonadal axis dysfunction; cardiovascular system;  
KW abnormal proliferative disorder; cancer; connective tissue dysfunction;  
KW epidermal system; elasticity.

OS Homo sapiens.

XX Key Location/Qualifiers

XX Protein I..127

XX FT /label= zalphal

XX FT /note= "mature protein fragment, no start codon given"

XX WO9929720-A2.

XX 17-JUN-1999.

XX 10-DEC-1998; 98WO-US26273.

XX 10-DEC-1997; 97US-0987926.

XX (ZYMO ) ZYMOGENETICS INC.

XX Conklin DC, Lok S, Parrish J;

XX WPI; 1999-385572/32.

XX N-PSDB; AAX80687.

XX Mammalian alpha helical protein-1, designated zalphal

XX Claim 5; Page 69; 73pp; English.

CC The present sequence represents the mature protein fragment of a novel  
CC alpha helical protein-1 designated zalphal from human pituitary gland  
CC cDNA library. Zalphal is predicted to be a four-helical protein similar  
CC to the family of helical cytokines represented by growth hormone,  
CC erythropoietin, leptin and interleukin-10. Zalphal gene was mapped to  
CC chromosome Xq27.3, in close proximity to FMR1, a gene linked to Fragile-X  
CC syndrome. Its transcription levels were found to be reduced or absent in  
CC Fragile-X patients. Zalphal transcripts were found at high levels in  
CC pituitary and aorta, and lower levels in brain, kidney, pancreas,  
CC prostate, etc. Zalphal is supposed to be responsible for the  
CC hypothalamic-pituitary-gonadal axis dysfunction and connective tissue  
CC dysfunctions in Fragile-X patients. The zalphal DNA is used in gene  
CC therapy for treating patients having a mutated zalphal gene or lacking  
CC the gene. Probes derived from zalphal gene can be used to check  
CC abnormalities on X chromosome. Zalphal protein may be useful in the  
CC treatment of Fragile-X syndrome and abnormal proliferative disorders e.g.  
CC cancer. It can also be used for the growth, differentiation, maintenance  
CC and survival of connective tissues, particularly cardiovascular and  
CC epidermal systems and in imparting cosmetic improvements to normal  
CC connective tissues such as enhancement of skin tone and elasticity.

XX Sequence 126 AA;

Query Match 7.2%; Score 155; DB 20; Length 126;

Best Local Similarity 27.6%; Pred. No. 6e-05;

Matches 40; Conservative 18; Mismatches 57; Indels 30; Gaps 2;

QY 102 ENNNLSQVOKLTENTTLREQVEPTPEDDDDIELRGAAAAA PPPPEECPEDLPEK 161

```
Db 2  ensqlmeqlrlilvceraallrqv-----pscpvpfpet 36
Qy 162  DGNPDMLAPFMAQCQIFMEKSTRDFSVDVRVCFVTSMTTGRAARWASAKLERSHYLMH 221
Db 37  fngessrlpefivqtasymlnenrfendankvafllilgtaeeewvvpviemdspllg 96
Qy 222  NYPAFMMEMKHVF-----EDPQRRE 241
Db 97  dyrafldemkqcfqgwdeddddee 121

RESULT 8
ID AAY26054 standard; Protein; 110 AA.
XX AC AAY26054;
XX KW Alpha helical protein-1; Fragile-x syndrome; cardiovascular system;
KW connective tissue; abnormal proliferative disorder; cancer; skin tone;
KW epidermal system; cosmetic improvement; skin tone; elasticity; Zalphal;
KW epitope.
XX OS Homo sapiens.
XX PN W09929720-A2.
XX PD 17-JUN-1999.
XX PF 10-DEC-1998; 98WO-US26273.
XX PR 10-DEC-1997; 97US-0987926.
XX PA (ZYMO ) ZYMOGENETICS INC.
XX PI Conklin DC, Lok S, Parrish J;
XX WPI; 1999-385572/32.
XX PT Mammalian alpha helical protein-1, designated zalphal
XX PS Claim 10; Page 70; 73pp; English.
XX CC The present sequence is an epitope-bearing protein derived from
CC human alpha helical protein-1, Zalphal. The protein comprises
CC helices A, B, C and D of mature Zalphal. It is used to raise specific
CC antibodies which can be used for detection and purification of Zalphal.
CC The Zalphal protein may be useful in the treatment of Fragile-X
CC syndrome and abnormal proliferative disorders e.g. cancer. It can also
CC be used for the growth, differentiation, maintenance and survival of
CC connective tissues, particularly cardiovascular and epidermal systems
CC and in imparting cosmetic improvements to normal connective tissues such
CC as enhancement of skin tone and elasticity.
XX SQ Sequence 110 AA;

Query Match 7.0%; Score 149.5; DB 20; Length 110;
Best Local Similarity 27.3%; Pred. No. 0.00014;
Matches 36; Conservative 18; Mismatches 53; Indels 25; Gaps 1;

Qy 103  NNNIQQVQKLTENTTLREQVEPTPEDEDDIELRGAAAAAAPPPIEECEPDLPKPF 162
Db 1  nsqimeqlrlilvceraallrqv-----pscpvpfpetf 35
Qy 163  DGNPDMLAPFMAQCQIFMEKSTRDFSVDVRVCFVTSMTTGRAARWASAKLERSHYLMH 222
Db 36  ngesrlpefivqtasymlnenrfendankvafllilgtaeeewvvpviemdspllg 95
```

```
Qy 223  YPAFMMEMKHVF 234
Db 96  yrafldemkqcf 107

RESULT 9
ID AAY94674 standard; Protein; 168 AA.
XX AC AAY94674;
XX DT 01-DEC-2000 (first entry)
XX DE Human zsig83 mature protein sequence.
XX KW Alpha-helical protein; zsig83; cell growth; differentiation; cancer;
KW proliferation; chromosome 22q13.1-ql3.2; cytostatic; vulneray;
KW degenerative condition; metastasis; wound healing.
XX OS Homo sapiens.
XX PH Key Location/Qualifiers
FT Domain 22..38
FT /note= "SH3-binding domain"
FT Region 41..55
FT /label= Helix_A
FT Region 60..65
FT /note= "Hydrophilic region"
FT Region 63..81
FT /label= Helix_B
FT Region 74..79
FT /note= "Hydrophilic region"
FT Region 97..105
FT /label= Helix_C
FT Region 103..108
FT /note= "Hydrophilic region"
FT Region 110..119
FT /label= Helix_D
FT Region 126..131
FT /note= "Hydrophilic region"
FT Region 128..138
FT /label= Helix_E
FT Region 162..167
FT /note= "Hydrophilic region"
XX PN W0200050594-A2.
XX PD 31-AUG-2000.
XX PF 25-FEB-2000; 2000WO-US04816.
XX PR 26-FEB-1999; 99US-0259131.
XX PA (ZYMO ) ZYMOGENETICS INC.
XX PI Presnell SR;
XX WPI; 2000-572091/53.
XX N-PSDB; AAA28032.
XX PT Alpha-helical protein zsig83, its antibodies and the polynucleotide
PT encoding the protein useful for treating disorders associated with
PT abnormal cell growth e.g. cancer and agonists useful for treating
PT wounds .
XX PS Claim 4; Page 75; 83pp; English.
XX CC This invention relates to a novel human alpha-helical protein designated
CC zsig83. Zsig83 plays a role in the process of cell growth,
CC differentiation, or proliferation. The zsig83 gene is located on
CC chromosome 22 at position 22q13.1-ql3.2. Included in the invention are
CC polynucleotide sequences encoding the zsig83 protein, expression vectors
CC containing the zsig83 DNA sequence, a cultured cell containing the
```







RESULT 13  
AAAY5927  
ID AAY5927 standard; Protein; 144 AA.  
XX  
AC AAY5927;  
XX  
DT 28-JAN-2000 (first entry)  
XX  
DE Human myometrium tumour EST encoded protein 7.  
XX  
KW Myometrium; tumour; human; expressed sequence tag; EST; uterine myoma;  
KW treatment; carcinoma; cancer; gene therapy.  
XX  
OS Homo sapiens.  
XX  
DEI9817947-A1.  
XX  
PD 28-OCT-1999.  
XX  
PF 17-APR-1998; 98DE-1017947.  
XX  
PR 17-APR-1998; 98DE-1017947.  
XX  
PA (META-) METAGEN GES GENOMFORSCHUNG MBH.  
XX  
PI Rosenthal A, Specht T, Hinzmann B, Schmitt A, Pilarsky C, Dahl E;  
DR WPI; 1999-602380/52.  
DR N-PSDB; AAZ41965.  
XX  
XX New nucleic acid sequences expressed in uterine myoma, and derived  
PT polypeptides, for treatment of uterine carcinoma and identification of  
PT therapeutic agents -  
XX  
PS Claim 23; Page 68; 86pp; German.  
XX  
XX This invention describes novel polypeptide sequences (I), fragments of  
CC (I) fragments and their encoding nucleic acids (II) which are highly  
CC expressed in human uterine myoma. (II) are used for recombinant  
CC expression of (I) and to isolate complete genes. (I) are used to  
CC identify agents suitable for treatment of uterine carcinoma, to directly  
CC treat this form of cancer (including expression from gene therapy  
CC vectors) and are used in a preparation for cancer treatment. (I) is also  
CC used for the generation of specific antibodies. (II) are identified by  
CC assembling ESTs (expressed sequence tags) from a particular tissue type  
CC before comparison of expression patterns. This allows a significantly  
CC longer fragment of the gene to be revealed and therefore reduces the  
CC number of failures associated with the fact that ESTs from different  
CC libraries may represent different parts of the same unknown gene,  
CC distorting the estimated frequency of occurrence in a particular tissue.  
CC AAY5921-Y59940 represent protein fragments encoded by the human  
CC myometrium tumour cDNA library derived EST fragments represented in  
CC AAZ41950-Z41980.  
XX  
XX Sequence 144 AA;

Query Match 6.4%; Score 137; DB 20; Length 144;  
Best Local Similarity 29.4%; Pred. No. 0.002;  
Matches 37; Conservative 19; Mismatches 58; Indels 12; Gaps 3;  
QY 121 REQVEPTPE-DEDDIELRGA-----AAAAAPPPPIEEEC--PEDLPKFDGNPDM 168  
Db 11 rrisadphatqtrnsaeargcmrdvqlmkallagplrpaarrwnlpfpetdgdtdr 70  
QY 169 LAPFMAQCQIPMEKSTRDPSVDRVRCFVTSMMTGTAARWASAKLERSHYLMHNPAPFMM 228  
Db 71 lpefivqtcsymfidentsndalkvfltrltgpalqwlpyirkespilndyrgfla 130  
QY 229 EMKHVF 234  
Db 131 emkrvf 136

RESULT 14  
AAW48895  
ID AAW48895 standard; Protein; 1230 AA.  
XX  
AC AAW48895;  
XX  
DT 13-OCT-1998 (first entry)  
XX  
DE Candida albicans CST20 protein.  
XX  
KW CST20; protein kinase; Ste20p/p65PAK family; screening; virulence;  
KW hyphal formation; pathogenic fungi; inhibitor; inflammation;  
KW antimycotic.  
XX  
OS Candida albicans.  
XX  
PN WO9818927-A1.  
XX  
PD 07-MAY-1998.  
XX  
PF 29-OCT-1997; 97WO-CA00809.  
XX  
PR 30-OCT-1996; 96US-0029458.  
XX  
PA (CANA ) NAT RES COUNCIL CANADA.  
XX  
PI Leberer E, Thomas DY;  
DR WPI; 1998-272222/24.  
DR N-PSDB; AAV32553.  
XX  
XX In vitro screening test for agents that inhibit Candida genes  
PT involved in virulence - and transition to hyphal form, potentially  
PT useful as antimycotic agents  
XX  
PS Disclosure; Fig 3; 79pp; English.  
XX  
XX The sequence is that of the CST20 protein which can be used  
CC in the development of an in vitro screening test for compounds  
CC that inhibit biological activity of the protein and a system for  
CC measuring its activity. The protein is involved in virulence and  
CC hyphal formation. Inhibitors are potentially useful for rendering  
CC pathogenic fungi (any species in which hyphal induction by kinase  
CC occurs) avirulent and/or to treat inflammation.  
XX  
XX Sequence 1230 AA;

Query Match 6.3%; Score 135.5; DB 19; Length 1230;  
Best Local Similarity 22.5%; Pred. No. 0.043;  
Matches 67; Conservative 54; Mismatches 114; Indels 63; Gaps 15;  
QY 17 GGQDPGLPHR-SEATAGRSPTTTLGPDPCPPPPPPPPNNNNNNNSKHTGHKSA--- 72  
Db 770 ggennalpkrinefahrapppp-plappappvpappanlliseqtseipqartaplq 827  
QY 73 -----CVPNMTERRRDELSEINNLRKVMKQSEENNLSQVOKLTFEINTTLREQVEP 126  
Db 828 aladvaptptniyeiqtkygeaqqklrekarelee-----lqrlreknergrqgqet 880  
QY 127 TPEDEDDIELRGAANAAPPPPIEEECPELPEKFDGNPD-MLAPFMAQ----- 175  
Db 881 gqnnad-----tasggsnialppvpvynkxp---psgs9ggrdakqaaliaqkrreekrkn 933  
QY 176 CQIFMEKST-----RDFSVDVRVRCFVTSMMTGTAARWASAKLERSHYLMHNPAPFMM 228  
Db 934 lqiaiktictncpdpnelyvdlvki-----ggasgvgvflahdvrdksnival 982  
QY 229 EMKHVFDPQREVAKRIRLRQGM-GSVTDYSNAPFQIAQDLWDNEPALIDYHEG 285  
Db 983 kqmnleqqp-kkeilneillvmkgslnpnlvfidsy-llkgdl-w-----vimeymeg 1033

[illegible]

Search completed: August 13, 2002, 15:11:11  
Job time: 242 sec

Query Match 6.28; Score 132; DB 19; Length 557;



**THIS PAGE BLANK (USPTO)**



; NUMBER OF SEQ ID NOS: 56  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 45  
; LENGTH: 126  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-209-525-45

Query Match 7.2%; Score 155; DB 4; Length 126;  
Best Local Similarity 27.6%; Pred. No. 6.2e-07;  
Matches 40; Conservative 18; Mismatches 57; Indels 30; Gaps 2;  
QY 102 ENNLSQVQKLTETNTLREQVEPTPEDEDDIELRGAAAAAAPPPIEECPEDLPK 161  
Db 2 ENSQLEQLRLVCRASLLRQVRP-----PSCVPFPET 36  
QY 162 FGNPDLAPFAQCOIFMEKSTRFSVDRVRCFVTSMTGTAARWASAKLERSHYLMH 221  
Db 37 FNGESSRLPEFTVQTASYMLVNRFCNDAMKVAFLISLLTGEAEWVYPIEMDSPILG 96  
QY 222 NYPAFMEKMHVF-----EDQRR 241  
Db 97 DYRAFIDEMKQCFGWDDDEDDDEE 121

RESULT 3  
US-09-209-525-52  
; Sequence 52, Application US/09209525  
; Patent No. 6303770  
; GENERAL INFORMATION:  
; APPLICANT: Lok, Si  
; APPLICANT: Conklin, Darrell C.  
; APPLICANT: Parrish, Julia E.  
; TITLE OF INVENTION: Mammalian Alpha Helical Protein-1  
; FILE REFERENCE: 97-71  
; CURRENT APPLICATION NUMBER: US/09/209,525  
; CURRENT FILING DATE: 1998-12-10  
; NUMBER OF SEQ ID NOS: 56  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 52  
; LENGTH: 110  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-209-525-52

Query Match 7.0%; Score 149.5; DB 4; Length 110;  
Best Local Similarity 27.3%; Pred. No. 1.6e-06;  
Matches 36; Conservative 18; Mismatches 53; Indels 25; Gaps 1;  
QY 103 NNLSQVQKLTETNTLREQVEPTPEDEDDIELRGAAAAAAPPPIEECPEDLPK 162  
Db 1 NSQLEQLRLVCRASLLRQVRP-----PSCVPFPET 35  
QY 163 DGNPDLAPFAQCOIFMEKSTRFSVDRVRCFVTSMTGTAARWASAKLERSHYLMH 222  
Db 36 NGESSRLPEFTVQTASYMLVNRFCNDAMKVAFLISLLTGEAEWVYPIEMDSPILG 95  
QY 223 NYPAFMEKMHVF 234  
Db 96 YRAFIDEMKQCF 107

RESULT 4  
US-08-574-959A-9  
; Sequence 9, Application US/08574959A  
; Patent No. 5962224  
; GENERAL INFORMATION:  
; APPLICANT: Jaekyoon Shin, Insil Joong, Ratna K. Vadlamudi  
; APPLICANT: and Jack L. Strominger  
; TITLE OF INVENTION: p62 POLYPEPTIDES, RELATED POLYPEPTIDES  
; TITLE OF INVENTION: AND USES THEREFOR

; NUMBER OF SEQUENCES: 22  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: LAHIVE & COCKFIELD  
; STREET: 60 State Street, Suite 510  
; CITY: Boston  
; STATE: Massachusetts  
; COUNTRY: USA  
; ZIP: 02109-1875  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/574,959A  
; FILING DATE: 19-DEC-95  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Mandragouras, Amy E.  
; REGISTRATION NUMBER: 36,207  
; REFERENCE/DOCKET NUMBER: DFN-008  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (617)227-7400  
; TELEFAX: (617)227-5941  
; INFORMATION FOR SEQ ID NO: 9:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 905 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-08-574-959A-9

Query Match 5.8%; Score 125; DB 2; Length 905;  
Best Local Similarity 20.5%; Pred. No. 0.0055;  
Matches 45; Conservative 22; Mismatches 75; Indels 78; Gaps 7;  
QY 22 GLPHRSEATAGRSPTPTVTTLGP-----DCPPPPPPPPPP----- 56  
Db 560 GLPPLPPPPSGATPP-PIAPTGPPTASPPVPAKEPEELPAAPGPLPPPPPPPPVPG 618  
QY 57 -----NNNNNNKHTGHKSACVNMTRRDE-- 84  
Db 619 VXLPPQLVPECTPGGGPPALEEDLTVININSSDEEEEEEDEEEDEEEDEEE 678  
QY 85 -----LSEINNREKVMKQSENNNLSQVQKLTETNTLREQVEPTPEDEDDIE-- 136  
Db 679 EEEEDFEEDEDEEYFEEEEEFEFEFEFEFEFEFEFEFEFEFEFEFEFEFEFEFE 735  
QY 137 -----LRGAAAAAPPPIEEEC--PEDLPKFDGNDP 167  
Db 736 EDLFGTAGGEVGEAGAPPPTLPPALPPPEPPKQVPEPE 775

RESULT 5  
US-09-357-014-9  
; Sequence 9, Application US/09357014  
; Patent No. 6291645  
; GENERAL INFORMATION:  
; APPLICANT: Jaekyoon Shin, Insil Joong, Ratna K. Vadlamudi  
; and Jack L. Strominger  
; TITLE OF INVENTION: p62 POLYPEPTIDES, RELATED POLYPEPTIDES  
; AND USES THEREFOR  
; NUMBER OF SEQUENCES: 22  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: LAHIVE & COCKFIELD  
; STREET: 60 State Street, Suite 510  
; CITY: Boston  
; STATE: Massachusetts  
; COUNTRY: USA  
; ZIP: 02109-1875  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible







Query Match 5.3%; Score 114; DB 3; Length 640;  
Best Local Similarity 19.4%; Pred. No. 0.034;  
Matches 64; Conservative 55; Mismatches 125; Indels 86; Gaps 13;



; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent In Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: PCT/US94/04496  
; FILING DATE:  
; CLASSIFICATION:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Deluca Esq., Mark  
; REGISTRATION NUMBER: 33,229  
; REFERENCE/DOCKET NUMBER: TJU-1242  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (215) 568-3100  
; TELEFAX: (215) 568-3439  
; INFORMATION FOR SEQ ID NO: 48:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1612 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; HYPOTHETICAL: NO  
; ANTI-SENSE: NO  
; PCT-US94-04496-48

Query Match 5.2%; Score 112; DB 5; Length 1612;  
Best Local Similarity 19.7%; Pred. No. 0.19;  
Matches 60; Conservative 41; Mismatches 138; Indels 66; Gaps 9;  
  
QY 31 TAGRSPTPTVTLGP---DCPPPPPPPPNNNNN-----NNSKHTGHSACV 74  
Db 1329 TPAATPATVAVSOPITDLPPTPPPPPVHYAGDFGMSMDLPLPPPSANQIGLPSAQV 1388  
  
QY 75 PNMTERRRDELSEINNREKVMKQSEENNLOSQVQKLTETNTTLREQVEPTPEDEDD 134  
Db 1389 AAERKRREHORWYE--KEKAPLEERERKRREQERKLGQMET---QSLNPAPFSPPLTA 1443  
  
QY 135 IELRGAAGAAAAPPP---TEEECPEDLPKFDGNPMDLAPFMAQCQIFMEKSTRDFSVD 191  
Db 1444 QMKPEKPTLQRPQETVIRELPQQPRTIE-----RRDLQYITVSK 1486  
  
QY 192 VRVCFVTSMTGRAARWASAKLERSHYLMHNPAPFMMEMKHVFEDPORREVAKRKIRRL 251  
Db 1487 EELSSGDSLSPPDPKRWKRAKEKLEKQOQ-MHIVDMSKEIQELQSKPDGSAEESDRLKLM 1545  
  
QY 252 QMGSGVIDYSNAPQMTAQDLDNNEPALIDQYHEGLSDHIQEELSHLEVAKSLSALIGQCI 311  
Db 1546 -----LEWQFQKRQESKQKDEDEDDDDVD-----TMLIMQRL 1581  
  
QY 312 HIERR 316  
Db 1582 EAERR 1586

RESULT 15  
US-09-370-368-9  
; Sequence 9, Application US/09370368  
; Patent No. 6258932  
; GENERAL INFORMATION:  
; APPLICANT: Anders Vahlne  
; TITLE OF INVENTION: PEPTIDES THAT BLOCK VIRAL INFECTIVITY  
; FILE REFERENCE: TRIPEP.003A  
; CURRENT APPLICATION NUMBER: US/09/370,368  
; CURRENT FILING DATE: 1999-08-09  
; NUMBER OF SEQ ID NOS: 10  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 9  
; LENGTH: 538  
; TYPE: PRT  
; ORGANISM: Moloney Murine Leukemia Virus  
US-09-370-368-9

Query Match 5.2%; Score 110.5; DB 4; Length 538;  
Best Local Similarity 19.0%; Pred. No. 0.057;  
Matches 96; Conservative 51; Mismatches 184; Indels 175; Gaps 22;  
  
QY 21 PGLHPHRS-----EATAGRSPTPTVTLIGPDCPPPPPPP-----PPNNNNN 61  
Db 71 PGPHGHPDQVPYIVTWEALAFDPPP---WVKPFVHPKPPPLLPSPAPSLPEPPLS--- 123  
  
QY 62 NNSKHTGHSACVPMNTERRRDELSEINNREKVMKQSEENNLOSQVQKLTETNTTLR 121  
Db 124 -----TPPQSSLYPALTPSL-----GAKPKQVLSDSGGPL-----IDLTEDPPPYR 166  
  
QY 122 EQVEPTPEDEDDDI-ELRGAAAAAP-----PPIEECPEDLPKFDGNPDM 168  
Db 167 DP-RPPPSDRDGSGEATPAGEAPDFSPMASRLGRREPPVADSTTSQAFPLRTGGNGQL 225  
  
QY 169 -LAPFMAQCQIFMEKSTRDFSVDVRV-CFVTSNMTGRAARW-----PPNNNNN 208  
Db 226 QYWPFSDDLNNKNNNPSFSDPGKLTALIESVLITHQPTWDDCQQLGLLTGEEKQR 285  
  
QY 209 -----ASAKLER-----SHYLMH----- 221  
Db 286 VLLEARKAVRGDDGRPTQLPNEVDAAFPLEPRPDWEYTTQAGRNHLVHYRQLLIAGLQNA 345  
  
QY 222 NYPAFMMEMKHVFEDPORREVAKRKIRRLRQMGSVIDYS-----NAPQMTAQDLDNNEPA 277  
Db 346 RSTNLAKEVIGTQGP--NESPSAFLEKAEYRRYTPYDPDPGQGTNVSMFSIMOSAP 403  
  
QY 278 LIDQYHEGLSDHIQEELSHLEVAKSLSALIGQCIHI-----ER 315  
Db 404 DIGRKLRL-----EDLRN-----KTLGDLVREAEIRFNKRETPEREERIRREERKEER 454  
  
QY 316 RLARAAAARKPRGPPRALVLPHTASHHVDPTPEPVGGARMRLTQEEKERRK---LNLCL 372  
Db 455 RRTEDQEKERDRRR-----HREMSRLLATVYSGQRDRQEGERRRRSOLDCDOCT 505  
  
QY 373 YCOTGGHYADNCPAKASKSKSPAGNSP 398  
Db 506 YCEEQGHWAQDCPR--PRGPRGPRP 529

Search completed: August 13, 2002, 15:11:34  
Job time: 240 sec

---

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 13, 2002, 15:08:59 ; Search time 20.9 Seconds  
(without alignments)  
1843.628 Million cell updates/sec

Title: US-09-631-863A-2  
Perfect score: 2139  
Sequence: 1 MRNKRVLTKKRRSGRGD.....DNCPAKASKSPAGNSAPPL 401

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283138 seqs, 96089334 residues

Total number of hits satisfying chosen parameters: 283138

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : PIR\_71.\*  
1: pir1.\*  
2: pir2.\*  
3: pir3.\*  
4: pir4.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB ID	Description
1	720.5	33.7	232	JE0163	myelin expression
2	240.5	11.2	853	S60178	gag polyprotein ho
3	190	8.9	457	T18347	gag protein homolo
4	159.5	7.5	639	S23569	gag polyprotein ho
5	148	6.9	1494	T13798	hypothetical prote
6	137	6.4	349	T18349	probable gag prote
7	135	6.3	537	F0M9RV	gag polyprotein -
8	134.5	6.3	555	T30349	structural protein
9	134	6.3	1529	A59310	unconventional myo
10	131.5	6.1	538	S70394	gag polyprotein -
11	131.5	6.1	1230	T18256	probable serine/th
12	131.5	6.1	1230	T18259	serine/threonine p
13	130	6.1	334	T01815	hypothetical prote
14	130	6.1	537	F0M9VB	gag polyprotein -
15	129	6.0	1110	T19673	hypothetical prote
16	129	6.0	1188	T46608	zinc finger protei
17	125	5.8	915	T26695	hypothetical prote
18	125	5.8	1651	T14160	transmembrane rece
19	123.5	5.8	1182	T30189	myelin transcripti
20	123.5	5.8	1585	T31611	hypothetical prote
21	122.5	5.7	536	F0M9ME	gag polyprotein -
22	122.5	5.7	538	S35474	gag polyprotein -
23	122.5	5.7	601	S33377	P63 protein - huma
24	122.5	5.7	1316	T00381	KIAA0633 protein -
25	121.5	5.7	1187	T46637	transcription fact
26	121.5	5.7	1585	T18274	1-phosphatidylinos
27	121	5.7	1612	T30805	du11 protein - mo
28	120	5.6	745	D96829	homeobox protein (
29	120	5.6	747	S71478	homeotic protein A

30	120	5.6	992	2	T46337	hypothetical prote
31	120	5.6	1905	2	T18267	multidrug resistan
32	120	5.6	1937	2	T18055	myosin heavy chain
33	119.5	5.6	1870	2	S37671	MHC class III hist
34	119.5	5.6	1872	2	S36152	MHC class III hist
35	119.5	5.6	2954	2	T14156	kinesin-related pr
36	119	5.6	320	2	AE2842	conserved hypotet
37	119	5.6	359	2	F97619	hypothetical prote
38	119	5.6	465	2	A02986	myosin alpha heavy
39	118.5	5.5	428	1	T36930	involucrin - white
40	118.5	5.5	7962	2	T38346	elastic titin - hu
41	118	5.5	2175	1	S03170	homeotic protein c
42	117	5.5	972	2	T49773	related to actin-i
43	116.5	5.4	249	2	A37280	C/EBP-related prot
44	116.5	5.4	708	2	T83196	NEDD-4 ORF - mouse
45	116.5	5.4	2142	2	B35098	MHC class III hist

ALIGNMENTS

RESULT 1

JE0163

myelin expression factor-3 - mouse

N:Alternate names: MYEF-3

C:Species: Mus musculus (house mouse)

C>Date: 03-Jul-1998 #sequence\_revision 10-Jul-1998 #text\_change 07-May-1999

C:Accession: JE0163

R:Stepilewski, A.; Krynska, B.; Tretiakova, A.; Haas, S.; Khalili, K.; Amini, S.

A:Title: MYEF-3, a developmentally controlled brain-derived nuclear protein which spe

A:Reference number: JE0163; MUID:98139908

A:Accession: JE0163

A:Molecule type: mRNA

A:Residues: 1-232 <STEP>

C:Experimental source: brain

C:Keywords: phosphoprotein

F:40-60/Domain: transmembrane #status predicted <TM>

F:31,175/Binding site: phosphate (Ser) (covalent) (by protein kinase C) #status predi

F:31,142,185,201/Binding site: phosphate (Ser) (covalent) (by casein kinase II) #stat

Query Match 33.7%; Score 720.5; DB 2; Length 232;

Best Local Similarity 61.6%; Pred. No. 1.6e-38;

Matches 149; Conservative 24; Mismatches 52; Indels 17; Gaps 5;

Qy 154 CPEDLPEKFDGNDMLAPFMAQCQIFMEKSTRDFSVDRVRCVFTSMGTGRA--ARMASA 211

Db 2 CLEDLPEKFDGNDMLGPFMYCQQLFMEKSTRDFSVDRIRVCVFTSMILIGRAPLGYCQAA 61

Qy 212 KL----ERSHYLMHNYPAFMEMKHVFEDPQRREVAKRKIRLRQGMGSVIDYSNAFOMI 267

Db 62 KMYPDAQLHCLYDGAELVLR-----PSESVSQTDQTSAPGPVGVVDYSNAFOMI 114

Qy 268 AQDLWDNEPALIDYHIEGLSQIEHLSHLEVAKSLSIGCQIHIERRLARAAAARPR 327

Db 115 AODLDWTEPALMDQFQELNPDIRAELSRQEAAPTALAIITACIHIERLRADAAA-RPD 173

Qy 328 SPPRALVLPHTASHHQVDPTPEVGGARMRLTQEEKERRKLNLCYCTGGHYADNCPAK 387

Db 174 PSPRALVMP---PNSQTDPTPEVGGARMRLSKEEKERRKNNLCYCGNGHGFADTCAK 230

Qy 388 AS 389

Db 231 AS 232

RESULT 2

S60178

gag polyprotein homolog - fungus (Fusarium oxysporum) retrotransposon skippy

C:Species: Fusarium oxysporum

C>Date: 15-Feb-1996 #sequence\_revision 01-Mar-1996 #text\_change 21-Jul-2000

C:Accession: S60178



C;Date: 13-Aug-1999 #sequence\_revision 13-Aug-1999 #text\_change 17-Nov-2000  
C;Accession: T13798  
R;Avedisov, S.  
submitted to the EMBL Data Library, February 1996  
A;Reference number: Z17761  
A;Accession: T13798  
A;Status: preliminary; translated from GB/EMBL/DBJ  
A;Molecule type: DNA  
A;Residues: 1-1494 <AVE>  
A;Cross-references: EMBL:X95908; NID:e990667; PID:e223896; PIDN:CAA65152.1  
C;Genetics:  
A;Cross-references: FlyBase:FBgn0002698  
A;Mobile element: retrotransposon mdg3

Query Match	6.9%	Score 148;	DB 2;	Length 1494;
Best Local Similarity	21.3%	Pred. No. 0.16;		
Matches 89;	Conservative	53;	Mismatches 155;	Indels 120;
Gaps				22;

Qy	39	PTVTGLGDCPP--PPPPPPPN--	-----NNNSKHTGHSACV	74
Db	41	PTAVRG--DCPEHPQKNAPGNDIFS	LDFQNCINTDHSVSNAMRKESTETG--	97
Qy	75	PNMTE--RRDELSER--INNLRKVMKQSE	-----NNNLSQVQK-----L	117
Db	98	TNMFELQOLRAELAEAKAMLNGTR	SLOFQEQOQBPQSKATVSSVQTQAO	157
Qy	118	TTLREQVEPTPEDEDDIELRGAAAAA	PPPTIEECPEDLPEKFGPMDLAPMAQ	177
Db	158	TTFSPORSNERAESQFPVDALALA	-----KETIDYDGT-----	199
Qy	178	IFMEKS--TRDFSVD--RVRVCFVTS	MMTGRAAKWASAKLERSHYLMHNY	229
Db	200	ITVVKNIARTFNTDDNHLRILLITK	--LKGNAQW-----LHAHPARLIE	247
Qy	230	--MKHVF--EDPORREVARRKIRRL	QMGSVTDYSNAFQIAQDLDWNEPALI	285
Db	248	LDQLSLTFGQSQAERIRRFESRKWK	TENFCGSYDEKMAUSNGINIDDELQ	307
Qy	286	LSDHIOELSHLEVAKSLSALIGQC	THIERRLARA--AAARKP--RSPPRAL	342
Db	308	IP--LQNFRTQARI-----QCSTP	SEMLRAFNSIRLPARREP-----	344
Qy	343	QVDPTEPVGGARMRLTQEE-----	KERRRKLNLCLYCGTGHYADNCPAK	392
Db	345	-VOPTYDKAIRCANCNSRGHKADIC	KPKRREPSCYACQGLHVAQCITRKS	400

RESULT 6  
T18349  
probable gag protein - rice blast fungus gypsy retroelement  
C:Species: Magnaporthe grisea (rice blast fungus)  
C:date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 04-Mar-2000  
C:Accession: T18349  
R:Robinson, K.F.  
submitted to the EMBL Data Library, September 1994  
A:Description: Sequence of the gpr retroelement.

A:Accession: T18349  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-349 <DO>  
A:Cross-references: EMBL:M77661; NID:g538065; PID:g538066; PIDN:AAA21441.1  
C:Genetics:  
C:Mobile element: gypsy retroelement  
C:Superfamily: rice blast fungus gypsy retroelement probable gag protein

Query Match	6.4%	Score 137;	DB 2;	Length 349;
Best Local Similarity	23.3%	Pred. No. 0.14;		
Matches 80;	Conservative	41;	Mismatches 124;	Indels 98;
				Gaps 16;

QY 105 NLOSVQKLTENTTLREQVEPTPEDEDDIELRGAAAAAPP--PPIEECPEDLPEKF 162

Db	27	DLOGRVOALQTGAPTVAIAE-----ALQATALPKRKPLRD-----PPLY	66
Qy	163	DGNPDLMLAPMAOCCIEMEXSTRDFSVDVRVRCVTSMTCTRAARWAS-----AKLERSH	217
Db	67	DGVP---ASFTAWKAMEYKLLROAD-----FIGDHQDQEYVLWAGLETSVQKVVRSY	116
Qy	218	YLM-----HNPYPAFMMEMKHVFEDPQORREVAKRKRIRLRQGMG-SVIDYSNAFQ---MI	267
Db	117	YEVGGRGAYRYTDFLDYLRERTYDDPHKRAQLAELETLMKPKQGSFAQFTAFERTLAT	176
Qy	268	AQDLWNNEPALIDQYHGLSDHIQEE-----LSHLEVAKSLSLALGQCIIH	312
Db	177	AGGLAWADEVRTNELFRVSPRIEACVGRGMGDTYLGAVAIYRQVAQDLEA-----IE	231
Qy	313	IERLA-----RAAAARKPRSPRALVLP--HTASHHQVDPTPEVGGARMR-----	356
Db	232	LDRFGFPHRAGAATAPR-PPKDEPTMTGTVAAM---GSRPNGGARGRRRRPGQTQPSDTN	286
Qy	357	-----LTOEEKERRRKLNLCLYCGTGGHYADNCPAKAS	389
Db	287	RRDTRPRAQWPSDEYQRRRETGACLRGNSGHQVADCTYAAA	329

RESULT 7

FOMVRV

gag polyprotein - radiation murine leukemia virus

N:Contains: core protein p15; core shell protein p30; inner coat protein p12; nucleop

C:Species: radiation murine leukemia virus

C:Date: 31-Mar-1989 #sequence\_revision 31-Mar-1989 #text\_change 16-Jul-1999

C:Accession: A26183

F:Merregaert, J.; Janowski, M.; Reddy, E.P.

Virology 158, 88-102, 1987

A:Title: Nucleotide sequence of a radiation leukemia virus genome.

A:Reference number: A94362; MUID: 87207680

A:Accession: A26183

A:Molecule type: DNA

A:Residues: 1-537 <MER>

A:Cross-references: GB:K03363; GB:M18449; NID:g332032; PTDN:AAA46518.1; PID:g332033

C:Genetics:

A:Gene: gag

C:Superfamily: mammalian retrovirus gag polyprotein I

C:Keywords: core protein; inner coat protein; nucleoprotein; polyprotein

F:1-129/Product: core protein p15 #status predicted <P15>

F:130-214/Product: inner coat protein p12 #status predicted <P12>

F:215-477/Product: core shell protein p30 #status predicted <P30>

F:478-537/Product: nucleoprotein p10 #status predicted <P10>

Query Match	6.3%	Score 135;	DB 1;	Length 537;
Best Local Similarity	19.1%	Pred. No. 0.32;		

Qy	21	PGLPHRS-----BATAGRSP-----TPVTTLGDCP-----PPPPPPPPNNNN	61
Db	71	PGPHGDQVPYIVTWEAIAIEPPSWKVPFVSPKLSLSTAPILPSGPSTQPP-----	124
Qy	62	NNSKHTGHKACVPNMTERRRDELSEINLRKVMKQSEENNLIQSOVKQLTENTTLR	121
Db	125	-----REALYPALTP-----SIKPPSPQVLSONGGLDILLTDEPPPYG	165
Qy	122	EQVEPTPEDEDDDIELRGAAAAAP-----PPPIEEBCPEDLPEKFDGNDPM	168
Db	166	EQGSPSPDGDREARYTSEIAPSPMVSRRLGRKRDPPAADSTTSRAFFLRLLGGNGQLQ	225
Qy	169	LAPMAOQCFIMEKSTRDEFSVDRVRV-CFTVSMTGRAARWASAKLERSHYLMHNPAPM	227
Db	226	YWFSSDDLWNKNNNSFSDPGKLTALTESVLTTHQPTWDDCQ-----L	272
Qy	228	MEMKHVEFDPQRREAVAKRKIRRLRQMGSVI-----DYSNAFQIAQDLWDNEP	276
Db	273	LGTLTTEGEORVLLREARKAVRGNDGRPTOLPNEVNSAFLEPRDMDYTTPEGRNHLVLY	332

QY 277 -----ALIDQYHEGL-----SDHIQF--- 292  
Db 333 ROLLLAGLQAGRSPTNLAKVIGITQGNESPSAFLERLKEAYRYTPYPEDHGQTSV 392  
QY 293 -----ELSHLE--VAKSLSAIGQCIHI-----E 314  
Db 393 SMSFIQSAFDIGRLKLERLEDKSLTLDLVREAEIFNKRTPPEERFRFRTEENEE 452  
QY 315 RRLARAAARKPSPPR---ALVPLHIAHQVDTPVGGAGMRLTQEKERRRKLNL 370  
Db 453 RRAEDEQREKERRRRQREMSKLLAVVTCQRODRO---GGERKR-PQLDKDQ----- 502  
QY 371 CLYCGTGCHVADNCPAKASKSSPAGNSP 398  
Db 503 CAYCKEGHAKDCPKK--PRGPRGPRP 528

RESULT 8  
T30349  
structural protein pp78-81 - Lymantria dispar nuclear polyhedrosis virus  
C:Species: Lymantria dispar nuclear polyhedrosis virus, LdMNPV  
C>Date: 29-Oct-1999 #sequence\_revision 29-Oct-1999 #text\_change 29-Oct-1999  
C:Accession: T30349  
R:Kuzio, J.; Pearson, M.N.; Harwood, S.H.; Funk, C.J.; Evans, J.T.; Slavicek, J.M.; Rohd  
Virology 253, 17-34, 1999  
A:Title: Sequence and analysis of the genome of a baculovirus pathogenic for Lymantria d  
A:Reference number: 220836; MUID:99124785  
A:Accession: T30349  
A:Status: preliminary; translated from GB/EMBL/DDBJ  
A:Molecule type: DNA  
A:Residues: 1-555 <KUZ>  
A:Cross-references: EMBL:AF081810; PIDN:AAC70187.1

Query Match 6.3%; Score 134.5; DB 2; Length 555;  
Best Local Similarity 20.7%; Pred. No. 0.35;  
Matches 75; Conservative 51; Mismatches 129; Indels 107; Gaps 16;

QY 21 PGLPHRSEATAGSP-----TPTVTGLGDCPPPPPPPPNNNNNNKHTGKSACV 74  
Db 214 PASEPARQESPIGSSAPEPIQETPTGLFAPPPPPPPPPPP-----PPEPQOKSSAV 267  
QY 75 -----PNMTERRRDELSEINNLRKVMKQSENNNLSQVOK-----LTEENT 118  
Db 268 PPPPPPLPPPGADDPFGEIGQEV---RPKPAERAPTALFAEIRRGVQLKPATERAP 324  
QY 119 T-----LRE--QVEPTPEDED-----DDIELRGAATAAAPP- 148  
Db 325 TYTPDALFAEIRQGVKLKPAARADEPPKPSRAPLLEIENRDKIKLKVAPRATPEPA 384  
QY 149 -----PIE-----ECPEDLPKFDGNPDLAPFMAQCQIFMEKSTRDFSV----- 190  
Db 385 SATNPLMQLLNKRLKESKKWSAESDAN-----YTSSWSDAEDDSLRLDA 431  
QY 191 -RVRVCFVTSMGTG-----AARWASAKLERSHYLMHNYPAFMMEMKHVPEDPORREVAK 244  
Db 432 LRIKALLGPRLSERSEKRIAKRLAGAKLSAEXTLDLQARATEPNDPLSPPYQLTAP 491  
QY 245 RKTRRLRQMGSVID-----YSNAFOMI--AQDLWNEPALIDQYHEGLSDHIQBELSH 296  
Db 492 LYHLDLKLFSAVLDLFRNGAYETALEKLEALQVDLQAPSL-QRMHDDISTYVYKQKR 550  
QY 297 LE 298  
Db 551 LE 552

RESULT 9  
A59310  
unconventional myosin heavy chain - maize  
N:Alternate names: MYO1  
C:Species: Zea mays (maize)  
C>Date: 09-Jun-2000 #sequence\_revision 09-Jun-2000 #text\_change 08-Sep-2000

C:Accession: A59310  
R:Liu, L.; Pesacreta, T.C.  
submitted to GenBank, May 1999  
A:Reference number: A59310  
A:Accession: A59310  
A:Status: preliminary; not compared with conceptual translation  
A:Molecule type: mRNA  
A:Residues: 1-1529 <LIU>  
A:Cross-references: GB:AF104924; NID:g4733890; PIDN:AAD17931.2; PID:g4733891  
C:Genetics:  
A:Gene: MYO1  
C:Superfamily: myosin MYO2; myosin motor domain homology  
F:65-719/Domain: myosin motor domain homology <MMO>

Query Match 6.3%; Score 134; DB 2; Length 1529;  
Best Local Similarity 22.3%; Pred. No. 1.2;  
Matches 73; Conservative 55; Mismatches 128; Indels 72; Gaps 14;

QY 78 TERRDELSEINNLRKVMKQSENNNLSQVOKLTEENTTLREQ---VEP----- 126  
Db 1005 SELNEELIKKFSAEKRIEQLTQVHRLEKATNMSENVKLRQQAIVASPTSKSLAAY 1064  
QY 127 -----TPEDDDDIELRGAATAAAPPPI-----EEECPEDLPEKFDGNPDML 169  
Db 1065 PKSPFQLKTPENG---ALNGEVKSSPDITPLPNPKLEAEERPKQSLNEKQEQENDLL 1121  
QY 170 APFMAQ-----CQIF-MEKSTRDFSDVRVCFVTSMMTGTAARWASAKLERS 216  
Db 1122 IKCVSQDLGFFSSGKPIAACLIYRCLLHWRSEFVERTGV-FDRIITQTGSAIESQDNNDKL 1180  
QY 217 HYLMHNPAPFMMEMKHVFE-----DPQREVAKRIRRLRQGM-GSVIDYSNAF---Q 265  
Db 1181 AYWLSNSTLLLLLQRTLKTTGAAGTFPQRRSSAAAFGRVFSGMRASPOGAFPMGSR 1240  
QY 266 MIA-----QDLWNEPAL-----IDQYHEGLSDHIQELSHLEVAKSLSALIGQCIHIER 315  
Db 1241 LIGGLGLDRQVEAKYPALLFKQLTAFLEKIYGMIRDNLK-----KEISPLGLGCIQAP- 1294  
QY 316 RLARAAARKPSPPRALVPLPHIASHHQ 343  
Db 1295 RTSRASLIKGRSQANALAQOQTLLIAHWQ 1322

RESULT 10  
S70394  
gag polyprotein - Friend murine leukemia virus (strain FB29)  
N:Contains: core protein p15; core shell protein p30; inner coat protein p12; nucleop  
C:Species: Friend murine leukemia virus  
A:Variety: strain FB29  
C>Date: 19-Mar-1997 #sequence\_revision 19-Mar-1997 #text\_change 20-Jun-2000  
C:Accession: S70394  
R:Perryman, S.; Nishio, J.; Chesebro, B.  
Nucleic Acids Res. 19, 6950, 1991  
A:Title: Complete nucleotide sequence of Friend murine leukemia virus, strain FB29.  
A:Reference number: S70393; MUID:92107687  
A:Accession: S70394  
A:Status: nucleic acid sequence not shown; translation not shown  
A:Molecule type: genomic RNA  
A:Residues: 1-538 <PBR>  
A:Cross-references: EMBL:Z11128; NID:g61547; PIDN:CAA77478.1; PID:g2654364  
A:Experimental source: strain FB29  
A>Note: the nucleotide sequence was submitted to the EMBL Data Library, October 1991  
C:Genetics:  
A:Gene: gag  
C:Superfamily: mammalian retrovirus gag polyprotein I  
C:Keywords: core protein; inner coat protein; nucleoprotein; polyprotein

Query Match 6.1%; Score 131.5; DB 2; Length 538;  
Best Local Similarity 18.8%; Pred. No. 0.53;  
Matches 95; Conservative 53; Mismatches 185; Indels 171; Gaps 17;









**THIS PAGE BLANK (USPTO)**

Result No.	Query			ID	Description
	Score	Match	Length		
1	159	7.4	146	LD01_HUMAN	O95751 homo sapien
2	156	7.3	520	CET1_CANAL	O93803 candida alb
3	135	6.3	537	GAG_MLVRD	P11269 radiation m
4	131.5	6.1	538	GAG_MLVFF	P26806 friend muri
5	131.5	6.1	1230	STZ0_CANAL	O92212 candida alb
6	130	6.1	537	GAG_MLVBM	P29167 murine leuk
7	129	6.0	872	S3B2_HUMAN	Q13435 homo sapien
8	122.5	5.7	536	GAG_MLVDE	P29168 murine leuk
9	122.5	5.7	538	GAG_MLVFP	P26805 dictyosteli
10	121.5	5.7	1585	P3K3_DICDI	P54675 dictyosteli
11	120.5	5.6	1157	Y182_HUMAN	Q14687 homo sapien
12	120.5	5.6	1816	AF6_HUMAN	P55196 homo sapien
13	120.5	5.6	1982	CHDM_DROME	O97159 drosophila
14	120	5.6	745	HGL2_ARATH	P46607 arabidopsis
15	120	5.6	1905	TAGB_DICDI	P54683 dictyosteli
16	120	5.6	1937	MYH8_HUMAN	P13535 homo sapien
17	119.5	5.6	1939	MYH1_HUMAN	P12882 homo sapien
18	119	5.6	465	MYH6_RABIT	P04460 oryctolagus
19	119	5.6	505	WASL_HUMAN	O00401 homo sapien
20	118.5	5.5	428	INVO_CEBAL	P24709 cebus albif
21	118.5	5.5	2564	SPOC_HUMAN	Q9h254 homo sapien
22	118	5.5	2175	HMCU_DROME	P10180 drosophila
23	117.5	5.5	501	WASL_RAT	O08816 rattus norv
24	116.5	5.4	281	CEBE_RAT	P56261 rattus norv
25	116.5	5.4	957	NED4_MOUSE	P46935 mus musculus
26	116.5	5.4	2142	BAT2_HUMAN	P48634 homo sapien
27	116	5.4	537	GAG_MLVAV	P03336 akv murine
28	116	5.4	3680	DMD_CANFA	O97592 canis famil
29	115.5	5.4	536	GAG_MLVCB	P27460 cas-br-e mu
30	115	5.4	1227	B3A3_MOUSE	P16283 mus musculus
31	115	5.4	1332	SPT7_YEAST	P35177 saccharomyc
32	115	5.4	1935	MYH7_HUMAN	P12883 homo sapien
33	115	5.4	1935	MYH7_PIG	P79293 sus scrofa



```
FT LIPID 2 2 MYRISTATE.
FT CONFLICT 479 T -> S (IN REF. 2).
FT CONFLICT 482 T -> S (IN REF. 2).
SQ SEQUENCE 537 AA: 60784 MW: 312AF7B2BB4B7FB CRC64;

Query Match
Best Local Similarity 6.3%; Score 135; DB 1; Length 537;
Matches 97; Conservative 58; Mismatches 173; Indels 180; Gaps 20;

QY 21 PGLHPRS-----EATAGSRP-----TPVTILGDCP-----PPPPPPNNNN 61
  || || || || || || || || || || || || || || || || || || || || ||
Db 71 PGPHGPDQVPIVITWEAIPSPMVSRRLGRKRDPAADSTTSRAFPRLGGNGQLQ 124
  : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 62 NNSKHTGHKSACVPNNTRRRDELSEINNLRKVMKQSENNNLSQVQKLTENTTLR 121
  : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 125 -----RSALYPALTP-----SIKPRSKPOVLSDNGGPLDLITEDPPPYG 165
  : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 122 EQVEPTPEDEDDIELRGAAAAAP-----PPPIEECPEDLPKFDGPNPDM- 168
  || || || || || || || || || || || || || || || || || || || || ||
Db 166 EQGSSPDGDDREATYTSIETAPSPMVSRRLGRKRDPAADSTTSRAFPRLGGNGQLQ 225
  : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 169 LAPFMAQOIFMEKSTRDSVDRVRV-CFVTSMTGAAARWASAKLERSHYLMHNPAPM 227
  || || || || || || || || || || || || || || || || || || || || ||
Db 226 YWPFSSDLYNWKNNPSEDPGKLTALIESVLTHQPTWDDCOQ-----L 272
  : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 228 MEMKHVFEDPQREYAKRIRLRQMGSVI--DYSNFMQIAQDLWNPE----- 276
  : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 273 LGTLTGEEKQVLLREAKRAGVNDGRPTQLPNEVNSAFPLERPDWYTTPEGRNHLVLY 332
  : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 277 -----ALIDQVHEGL-----SDHIQEQ----- 292
  || || || || || || || || || || || || || || || || || || || || ||
Db 333 RQLLAGLQAGRSPTNLAKVKITGTPNESPFAFLERLKEAYRYTPYDPEDHGQETSV 392
  : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 293 -----ELSHLE--VAKLSALIGQCIIH-----E 314
  : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 393 SMSFIWQSAPIGRKLERLEDLSKTLRLDVREAEIFNKRTPPEERERFRRETEENEE 452
  : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 315 RLRLAAAAKPRSPRP-----ALVLPHIASHHQVDPTPEVGGARMRLTOEERERKLN 370
  || || || || || || || || || || || || || || || || || || || || ||
Db 453 RRAEDEQEKERDRRRQREMSKLLATVYTGQRDRQ---GGERK-PQLDKDQ----- 502
  : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 371 CLYCGTGHYADNCPAKASKSPAGNSP 398
  || || || || || || || || || || || || || || || || || || || || ||
Db 503 CAYCKEGHWAKDCPKK--PRGPRGPRP 528
  : : : : : : : : : : : : : : : : : : : : : : : : : : : :

RESULT 4
GAG_MLVFF
ID GAG_MLVFF STANDARD; PRT; 538 AA.
AC P26806;
DT 01-AUG-1992 (Rel. 23, Created)
DT 01-AUG-1992 (Rel. 23, Last sequence update)
DT 01-MAR-2002 (Rel. 41, Last annotation update)
DE GAG polyprotein (core polyprotein) [Contains: Matrix protein P15; RNA
DE binding phosphoprotein P12; Capsid protein P30; Nucleocapsid protein
DE P10].
DE GAG.
GN Friend murine leukemia virus (isolate FB29) (F-MuLV).
OS Viruses; Retroviral viruses; Retroviridae; Mammalian type C retroviruses.
OC NCBI_TaxID=11797;
OX [1]
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=92107687; PubMed=1762923;
RT Perryman S., Nishio J., Chesebro B.;
RA "Complete nucleotide sequence of Friend murine leukemia virus, strain
RL FB29."
RT Nucleic Acids Res. 19:6950-6950(1991).
RN [2]
RP SEQUENCE OF 479-512.
RX MEDLINE=81264245; PubMed=6267042;
RA Henderson L.E., Copeland T.D., Sowder R.C., Smythers G.W.,
RA Oroszlan S.;
```

```
RT "Primary structure of the low molecular weight nucleic acid-binding
RT proteins of murine leukemia viruses.";
RL J. Biol. Chem. 256:8400-8406(1981).
CC -|- P1M: SPECIFIC ENZYMIC CLEAVAGES IN VIVO YIELD MATURE PROTEINS.
CC -|- MISCELLANEOUS: THIS PROTEIN IS SYNTHESIZED AS A GAG-POL
CC POLYPROTEIN.
CC -|- SIMILARITY: CONTAINS 1 CCHC-TYPE ZINC FINGER.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; Z11128; CAA77478.1; -.
DR InterPro; IPR000840; Gag_MA.
DR InterPro; IPR002079; Gag_P12.
DR InterPro; IPR003036; Gag_P30.
DR InterPro; IPR001878; znf_CCHC.
DR Pfam; PF01140; gag_MA; 1.
DR Pfam; PF01141; gag_P12; 1.
DR Pfam; PF02093; Gag_P30; 1.
DR Pfam; PF00098; znf_CCHC; 1.
DR SMART; SM00343; znf_C2HC; 1.
DR PROSITE; PS50158; ZF_CCHC; 1.
KW Coat protein; Core protein; Polyprotein; Nucleoprotein; Myristate;
KW Phosphorylation; Zinc-finger.
FT CHAIN 2 131 MATRIX PROTEIN P15.
  FT CHAIN 132 215 RNA BINDING PHOSPHOPROTEIN P12.
  FT CHAIN 216 478 CAPSID PROTEIN P30.
  FT CHAIN 479 538 NUCLEOCAPSID PROTEIN P10.
  FT ZN_FING 502 519 CCHC-TYPE.
  FT LIPID 2 2 MYRISTATE (BY SIMILARITY).
  FT CONFLICT 501 501 H -> R (IN REF. 2).
  SQ SEQUENCE 538 AA: 60929 MW: 2652DDB9E0D4A3C CRC64;

Query Match
Best Local Similarity 6.1%; Score 131.5; DB 1; Length 538;
Matches 95; Conservative 53; Mismatches 185; Indels 171; Gaps 17;

QY 21 PGLHPRS-----EATAGSRP-----TPVTILGDCP-----PPPPPPNNNNNNNS 64
  || || || || || || || || || || || || || || || || || || || || ||
Db 71 PGPHGPDQVPIVITWEAIPSPMVSRRLGRKRDPAADSTTSRAFPRLGGNGQYWP 123
  : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 65 KHTGHKSACVPNNTRRRDELSEINNLRKVMKQSENNNLSQVQKLTENTTLREQV 124
  || || || || || || || || || || || || || || || || || || || || ||
Db 124 --TPQSSLYPALT-----SPLNTKPRQVLPDSG-----GPLIDLITEDPPPYRDPG 169
  : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 125 EPTPEDEDDIELRGAAAAAP-----PPPIEECPEDLPKFDGPNP-MLAP 171
  || || || || || || || || || || || || || || || || || || || || ||
Db 170 PPSPDGNGDSGEVAPTEGAPDFSPMVSRRLGRKRDPAADSTTSRAFPRLGGNGQYWP 239
  : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 172 FNAQCQIFMEKSTRDSVDRVRV-CFVTSMTGAAARWASAKLERSHYLMHNPAPM 230
  || || || || || || || || || || || || || || || || || || || || ||
Db 230 FSSSDLYNWKNNPSEDPGKLTALIESVLTHQPTWDDCOQ-----LLGT 276
  : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 231 KHVFEDPQREYAKRIRLRQMGSVI--DYSNFMQIAQDLWN----- 274
  : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 277 LLTGEEKQVLLREAKRAGVNDGRPTQLPNDINAPLERPDWYNTQGRNHLVHYRQL 336
  : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 275 -----EPALIDQVHEGL----- 286
  : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 337 LLAGLQAGRSPTNLAKVKITGTPNESPFAFLERLKEAYRYTPYDPEDPQGTNVAMS 396
  : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 287 -----SDHIQEELSHLE--VAKLSALIGQCIIH-----ERRL 317
  : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 397 FIWQSAPIGRKLERLEDLSKTLRLDVREAEIFNKRTPPEERERIRRETEEKERR 456
  : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 318 ARAAAAARKPRSPRALVPLHSHHOVDPTPEVGGARMRLTOEERERKRL---NLCLYC 374
  || || || || || || || || || || || || || || || || || || || || ||
```

Db	457	AEDVQREKERDRRR-----HREMSKLLATVVVSGQRDQGGERRRRQLDHDQCAVC 507	
Qy	375	GTGGHYADNCAPAKASKSPAGNSP 398	
Db	508	KEKGHWARDCPKK--PRGPRGPRP 529	
RESULT	5		
ST20_CANAL			
ID	ST20_CANAL	STANDARD; PRT; 1230 AA.	
AC	Q92212;		
DT	01-NOV-1997 (Rel. 35, Created)		
DT	01-NOV-1997 (Rel. 35, Last sequence update)		
DT	16-OCT-2001 (Rel. 40, Last annotation update)		
DE	Serine/threonine-protein kinase STE20 homolog (EC 2.7.1.1-).		
GN	HST20 OR GST20.		
OS	Candida albicans (Yeast).		
OC	Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;		
OC	Saccharomycetales; mitosporic Saccharomycetales; Candida.		
OX	NCBI_TaxID=5476;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN=1066;		
RX	MEDLINE=97075146; PubMed=8917572;		
RA	Kohler J.R., Fink G.R.;		
RT	"Candida albicans strains heterozygous and homozygous for mutations		
RT	in mitogen-activated protein kinase signaling components have defects		
RT	in hyphal development."		
RL	Proc. Natl. Acad. Sci. U.S.A. 93:13223-13228(1996).		
CC	-!- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).		
CC	-!- SIMILARITY: BELONGS TO THE SER/THR FAMILY OF PROTEIN KINASES.		
CC	STE20 SUBFAMILY.		
CC	-!- SIMILARITY: CONTAINS 1 GBD DOMAIN.		
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration		
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -		
CC	the European Bioinformatics Institute. There are no restrictions on its		
CC	use by non-profit institutions as long as its content is in no way		
CC	modified and this statement is not removed. Usage by and for commercial		
CC	entities requires a license agreement (See <a href="http://www.isb-sib.ch/announcement/">http://www.isb-sib.ch/announcement/</a>		
CC	or send an email to <a href="mailto:license@isb-sib.ch">license@isb-sib.ch</a> ).		
DR	EMBL; U73457; AAB38875.1; -		
DR	HSSP; P00518; 1PKH.		
DR	InterPro; IPR000719; Euk_pkinase.		
DR	InterPro; IPR000095; PAK_Box_P21_Rho_bindng.		
DR	InterPro; IPR002290; Ser_thr_pkinase.		
DR	Pfam; PF00786; PBD; 1.		
DR	Pfam; PF00069; pkinase; 1.		
DR	SMART; SM00285; PBD; 1.		
DR	SMART; SM00220; S_TKC; 1.		
DR	PROSITE; PS50108; GBD; 1.		
DR	PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.		
DR	PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.		
DR	PROSITE; PS00108; PROTEIN_KINASE_ST; 1.		
KW	Transferase; Serine/threonine-protein kinase; ATP-binding.		
FT	DOMAIN 109 119 POLY-ASN.		
FT	DOMAIN 290 293 POLY-PRO.		
FT	DOMAIN 453 464 POLY-SER.		
FT	DOMAIN 552 555 POLY-SER.		
FT	DOMAIN 573 580 POLY-GLY.		
FT	DOMAIN 727 732 POLY-PRO.		
FT	DOMAIN 790 794 POLY-PRO.		
FT	DOMAIN 475 532 GBD.		
FT	DOMAIN 953 1205 PROTEIN_KINASE.		
FT	NP_BIND 959 967 ATP (BY SIMILARITY).		
FT	BINDING 983 988 ATP (BY SIMILARITY).		
FT	ACT_SITE 1073 1073 BY SIMILARITY.		
SQ	SEQUENCE 1230 AA; 132862 MW; 2B2AC4C133B9FE81 CRC64;		
Query Match	6.18; Score 131.5; DB 1; Length 1230;		
Best Local Similarity	21.5%; Pred. No. 0.46;		

Matches	64; Conservative	51; Mismatches	121; Indels	61; Gaps	13;
Qy	17	GGDPLGPHR-SEATAGRSPTPTVTTLGDCPPPPPPPPNNNNNNNSKHTGKSA---	72		
Db	770	GGENNALPQRINEFKAHRAPPP--PSAPPAPPVPPAPPANLLSEQTSEIPQQTAFSQ 827			
Qy	73	-----CVPNMTRRRRDEEINNLRKVMKQSENNNLSQVOKLTETTELREQVEP 126			
Db	828	ALADVTAPTNIYEIQTKYQEAQKLRKKARELEE-----IQRLEKRENRQROQET 880			
Qy	127	TPEDEDDIELRGAAAAAAPPPIEECPEDLPKFDGNPD-MLAPFMAQ----- 175			
Db	881	GQNAD-----TASGSNIAAPPVPVFNKKP---PSGGGGGRDAQAALIAQKREKKRKN 933			
Qy	176	COIFMEKST-----RDFSVDVRVCFVTSMTTGAARWASAKLERSHYLMHNPAPFMM 228			
Db	934	LQITAKLTICNPGDPNELYDLVKI-----CGGASGGVFLAHDVRDKSNIVAI 982			
Qy	229	EMKHVFEDPQRREYAKRKIRLRQGMGSVIDYSNAFQIAQDLWNEPALIDQYHEG 285			
Db	983	KOMNLEQPKKELIINEILVMKGGSSHPNIVNFIDSY-LLKGDL-W---VIMEYMEG 1033			
RESULT	6				
GAG_MLVB					
ID	GAG_MLVB	STANDARD; PRT; 537 AA.			
AC	P29167;				
DT	01-DEC-1992 (Rel. 24, Created)				
DT	01-DEC-1992 (Rel. 24, Last sequence update)				
DT	01-MAR-2002 (Rel. 41, Last annotation update)				
DE	GAG polyprotein [Contains: Core protein p15; Inner coat protein p12;				
DE	Core shell protein p30; Nucleoprotein p10].				
GN	GAG.				
OS	Murine leukemia virus (strain BM5 eco).				
OC	Viruses; Retroviridae; Retroviridae; Gammaretrovirus.				
OX	NCBI_TaxID=31687;				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RX	MEDLINE=91303677; PubMed=1649328;				
RA	Chattopadhyay S.K., Sengupta D.N., Fredrickson T.N., Morse H.C. III,				
RA	Hartley J.W.;				
RT	"Characteristics and contributions of defective, ecotropic, and mink				
RT	cell focus-inducing viruses involved in a retrovirus-induced				
RT	immunodeficiency syndrome of mice."				
RL	J. Virol. 65:4232-4241(1991).				
CC	-!- PTM: SPECIFIC ENZYMATIC CLEAVAGES IN VIVO YIELD MATURE PROTEINS.				
CC	-!- SIMILARITY: CONTAINS 1 CCHC-TYPE ZINC FINGER.				
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration				
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -				
CC	the European Bioinformatics Institute. There are no restrictions on its				
CC	use by non-profit institutions as long as its content is in no way				
CC	modified and this statement is not removed. Usage by and for commercial				
CC	entities requires a license agreement (See <a href="http://www.isb-sib.ch/announcement/">http://www.isb-sib.ch/announcement/</a>				
CC	or send an email to <a href="mailto:license@isb-sib.ch">license@isb-sib.ch</a> ).				
DR	EMBL; M64095; AAA46510.1; -				
DR	PIR; A40416; FOMVMB.				
DR	InterPro; IPR000840; Gag_MA.				
DR	InterPro; IPR002079; Gag_p12.				
DR	InterPro; IPR003036; Gag_p30.				
DR	InterPro; IPR001878; Znf_CCHC.				
DR	Pfam; PF01140; gag_MA; 1.				
DR	Pfam; PF01141; gag_p12; 1.				
DR	Pfam; PF02093; Gag_p30; 1.				
DR	Pfam; PF00098; zf-CCHC; 1.				
DR	SMART; SM00343; Znf_C2HC; 1.				
DR	PROSITE; PS0158; Zf_CCHC; 1.				
KW	Coat protein; Core protein; Polyprotein; Nucleoprotein; Myristate;				
KW	Zinc-finger.				
FT	CHAIN 2 129 CORE PROTEIN p15.				
FT	CHAIN 130 214 INNER COAT PROTEIN p12.				
FT	CHAIN 215 477 CORE SHELL PROTEIN p30.				



```
FT CHAIN 478 537 NUCLEOPROTEIN P10.
FT ZN_FING 501 518 CCHC-TYPE.
FT LIPID 2 MYRISTATE (BY SIMILARITY).
SQ SEQUENCE 537 AA; 60422 MW; ABD2E70299BFED64 CRC64;

Query Match 6.1%; Score 130; DB 1; Length 537;
Best Local Similarity 18.5%; Pred. No. 0.22;
Matches 93; Conservative 53; Mismatches 186; Indels 172; Gaps 17;

QY 21 PGLHPRS-----EATGRSPP-----TPVTVLGDCP-----PPPPPPNNNN 61
   || : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 71 PGPHGHDPQVYIVTWEATAYEPWPVFPVSKLSPTAILSPGSPSTOPP----- 124
   || : | : | : | : | : | : | : | : | : | : | : | : | : | : |
QY 62 NNSKHTGHSACVPNMTERRRDELSEENLNRLKVMKQSENNNLSQVOKLTTEINTLR 121
   || : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 125 -----RSALYPAFT-----PSIKRPSKQVLSDDGGPL---IDLUTEPPPYG 165
   || : | : | : | : | : | : | : | : | : | : | : | : | : | : |
QY 122 EQVEPTDEDDIELRGAATAAP-----PPPIEECEPDLPEKFDGNPDW- 168
   || : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 166 EQGPSSPDGDDREATSTSEIPAPSPMVSRLLRGKRDPADSTTSRAPPLRLGCGQLQ 225
   || : | : | : | : | : | : | : | : | : | : | : | : | : | : |
QY 169 LAPFMAQCOIFMEKTRDFSDVRVY-CFVTSMTGRAARWASAKLERSHYLMHNPAPM 227
   || : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 226 WYFSSSDLYNWKNNPFSFSDPGKLTALIESVLTHTOPTWDDCOQ-----L 272
   || : | : | : | : | : | : | : | : | : | : | : | : | : | : |
QY 228 MEMKHVEEDPQREVAKRIRLRGMSVI--DYSNAFQMLAQDLDMNEP----- 276
   || : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 273 LGTLTGEEKQVLEAKRVRGNDGRPTQLPNEVNSAPFLRPDWDYTPBGRNHLVLY 332
   || : | : | : | : | : | : | : | : | : | : | : | : | : | : |
QY 277 -----ALIDQVHEGLSDH----- 289
   || : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 333 RQLLAGLQAGRSPTNLAKYKITQGNPSFSAFLERLKEARYRTPYDPEDPQETNV 392
   || : | : | : | : | : | : | : | : | : | : | : | : | : | : |
QY 290 -----IQELSHLE--VAKLSALIGQCCHI-----E 314
   || : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 393 SMSFIWQSAAPAGRLERLEDLKSITGLDVREAEIFNKRETPEREERIRRETEKEE 452
   || : | : | : | : | : | : | : | : | : | : | : | : | : | : |
QY 315 RRLAARAARPRSPRALVLPFHIAHQVDPTPEVGGARMRLTOEERERRKLNLCLYC 374
   || : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 453 RRRAGDEQREKRRRRRREMSKLIATVVTGQRQRRGRRRQLDKDQ-----CAYC 506
   || : | : | : | : | : | : | : | : | : | : | : | : | : | : |
QY 375 CTGGHYADNCAPAKSKSPAGNSP 398
   || : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 507 KEKGHWKDCPKK--PRGPRGPRP 528

RESULT 7
S3B2_HUMAN
ID S3B2_HUMAN STANDARD; PRT; 872 AA.
AC Q13435;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Splicing factor 3B subunit 2 (Spliceosome associated protein 145) (SAP
DE 145) (SF3b150) (Pre-mRNA splicing factor SF3b 145 kDa subunit).
GN SF3B2 OR SAP145
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N. A. AND SEQUENCE OF 151-159 AND 794-817.
RX MEDLINE=96154048; PubMed=8566756;
RA Gozani O., Feld R., Reed R.;
RT "Evidence that sequence-independent binding of highly conserved U2
RT snRNP proteins upstream of the branch site is required for assembly
RT of spliceosomal complex A."
RL Genes Dev. 10:233-243(1996).
RN [2]
RP CHARACTERIZATION OF THE SPLICOSOME.
RX MEDLINE=20337962; PubMed=10882114;
RA Das R., Zhou Z., Reed R.;
```

```
RT *Functional association of U2 snRNP with the ATP-independent
RT spliceosomal complex E."
RL Mol. Cell 5:779-787(2000).
CC -!- FUNCTION: SUBUNIT OF THE SPLICING FACTOR SF3B REQUIRED FOR 'A'
CC COMPLEX ASSEMBLY FORMED BY THE STABLE BINDING OF U2 snRNP TO THE
CC BRANCHPOINT SEQUENCE (BPS) IN PRE-mRNA. SEQUENCE INDEPENDENT
CC BINDING OF SF3A/SF3B COMPLEX UPSTREAM OF THE BRANCH SITE IS
CC ESSENTIAL, IT MAY ANCHOR U2 snRNP TO THE PRE-mRNA. MAY ALSO BE
CC INVOLVED IN THE ASSEMBLY OF THE 'E' COMPLEX. BELONGS ALSO TO THE
CC MINOR U12-DEPENDENT SPLICOSOME, WHICH IS INVOLVED IN THE SPLICING
CC OF RARE CLASS OF NUCLEAR PRE-mRNA INTRON.
CC -!- SUBUNIT: COMPONENT OF SPLICING FACTOR SF3B WHICH IS COMPOSED OF
CC FOUR SUBUNITS; SF3B4/SAP49, SF3B3/SAP130, SF3B2/SAP145,
CC SF3B1/SAP155. SF3B ASSOCIATES WITH THE SPLICING FACTOR SF3A AND A
CC 12S RNA UNIT TO FORM THE U2 SMALL NUCLEAR RIBONUCLEOPROTEINS
CC COMPLEX (U2 snRNP). SF3B2 INTERACTS DIRECTLY WITH SF3B4.
CC -!- SUBCELLULAR LOCATION: NUCLEAR (BY SIMILARITY).
CC -!- SIMILARITY: TO YEAST CUS1. SOME, TO C.ELEGANS ZK632.11.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; U41371; AAA97461.1; -.
CC MIM; 605591; -.
CC DR InterPro; IPR003034; SAP.
CC DR Pfam; PF02037; SAP; 1.
CC DR SMART; SM00513; SAP; 1.
CC KW Spliceosome; mRNA processing; mRNA splicing; Nuclear protein.
CC FT DOMAIN 58 73 POLY-PRO.
CC FT DOMAIN 81 90 POLY-PRO.
CC FT DOMAIN 106 109 POLY-PRO.
CC FT DOMAIN 226 230 POLY-PRO.
CC FT DOMAIN 269 274 POLY-GLU.
CC FT DOMAIN 308 312 POLY-LYS.
CC FT DOMAIN 676 679 POLY-GLU.
CC FT DOMAIN 697 703 POLY-GLU.
CC SQ SEQUENCE 872 AA; 97656 MW; AED69FDD0DA5DE31 CRC64;

Query Match 6.0%; Score 129; DB 1; Length 872;
Best Local Similarity 20.7%; Pred. No. 0.45;
Matches 83; Conservative 64; Mismatches 150; Indels 104; Gaps 19;

QY 16 RGGQDPG-----LPHRSEATAGRSPTPTV-TLGPDCPPP--PPPPPPNNNNNNNSK 65
   | || | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 182 RPPQDMGQIGVFTPLGP-RVAPVGVGPTTVLPKGAVPVPRGPPPPGDNREMDDP 240
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 66 HTGHSACVPNMTER---RRDELSEENLNRLKVMKQSENNNLSQVOKLTTEINTLRE 122
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 241 SVGPK--IPOALEKITLQKSRQEMNSQEEEMETDARSLSGASASETEDTVSVSK 297
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 123 Q-----VEPTPEDEDDIELRGAATAAPPPPIEECEPDLPE 160
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 298 KEKNRKRNRKKKKKQPRVGVSSSSGDRKSDTSRSGSDSPA--DVEIYVTEEP 354
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 161 KPDGN-----PDLAPFMAQCQIFMEKSTRDFSVDRVRVCFVTSMTGRAARWASAKLERS 216
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 355 IYEPNFIFPKRIFEAFKLTDDVKKEKEPEKLD-----KLENS 393
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 217 HYLMHNPAPFMEMKHVPED-----PQREVAKRIRLRQMGSGVIDYSNAFQMI 267
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 394 --AAPKKGKFEHEKSDSDSDDEQEKKEAPKLSKKLRRMR-----FTVAELQLV 446
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 268 AQDLDMNEPALIDQYHEGLSDHIEQLSHLEVAKLSALIGQCIIHIERLARAAAAARKPR 327
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 447 AR-----PDVVEMHDVTAQD--PKLLVHLKATRN-SVPVPRHWCCKRKYLO--GKRGIE 495
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 328 SPPRALVLPFHIAHQVDPTPEVGGARMRLTOEERERRKL 368
```



DR PROSITE; PS50158; 2F\_CCHC; 1.  
KW Coat protein; Core protein; Polyprotein; Nucleoprotein; Myristate;  
FT Phosphorylation; Zinc-finger.  
FT CHAIN 2 131 MATRIX PROTEIN P15  
FT CHAIN 132 215 RNA BINDING PHOSPHOPROTEIN P12.  
FT CHAIN 216 478 CAPSID PROTEIN P30.  
FT CHAIN 479 538 NUCLEOCAPSID PROTEIN P10.  
FT ZN\_FING 502 519 CCHC-TYPE.  
FT LIPID 2 2 MYRISTATE (BY SIMILARITY).  
SQ SEQUENCE 538 AA; 61033 MW; 2F9F97D2C79DEBE CRC64;

Query Match 5.7%; Score 122.5; DB 1; Length 538;  
Best Local Similarity 18.6%; Pred. No. 0.65;  
Matches 98; Conservative 49; Mismatches 164; Indels 215; Gaps 20;

QY 21 PGLPHRS-----EATGRSP-----PTPVITLGPDCPPPPPPPPPPNNNNNS 64  
DB 71 PGPHGPDQVYIVTWELAVDPWVKPFVHPKPLLLPPSAPSLPEPPLS-----123  
QY 65 KHTGHSACVPNNMTRRRDELSEIINNLEKVMQSEENNLSQVQKLTENTTLREQV 124  
DB 124 --TPPOSSLYPALT-----SPLTKPRQVLPDSG-----GPLDLITEDPPPYRDPG 169  
QY 125 EPTPEDEDDIELRGAAAAAP-----PPPIEECEPEDLPKFDGNDP-ML 169  
DB 170 PPSPDCKGDSGEV--APTEGAPDSSPMVSLRGRREPPVADSTTSQAFLRLGGNGQFOY 227  
QY 170 APMAQCQIFMEKSTDFSVDRVRV-CFTVSMGTGRAAKWASAKLERSHYLMHNYPAFMM 228  
DB 228 WPFSSDLYNWKNNNPSFSDPKLTALIESVLLTHQPTWDDCQ-----LL 274  
QY 229 EMKHVEDPQRREVAKRRLRROGMSVI--DYSNAFOMIAQDLWDNEPALIDQYHEGL 286  
DB 275 GTLLTGEKQVRLLEARKAVRGEDGPTQLPNDINDAPFLERPDWDYN-----322  
QY 287 SHIQEELSHLEVAKSLALIGQCIHIERLARAAAARKPRSPRALVLPFIAS-----340  
DB 323 ---TORGRNHL-----VHY-RQLLAGLQNAGRSPTNLAKVKGITQGNESP 365  
QY 341 -----HHQVDTEP-----VGGARMRL-----357  
DB 366 SAFLEKLEAYRYTYPDPEDPQETNVSMFTWQSAPDIGRKLERLEDLKNKTLGLVLR 425  
QY 358 -----TOEEKERR-----KL-----368  
DB 426 EAEKFNKRETPEREERVRRETEKEERRRADERREKERDRRRHREMSKLLATVVSQG 495  
QY 369 -----NLCLYCGTGCHYADNCPAKASKSPAGNSP 398  
DB 486 RQDRGGERRRPQLDHDQCAYPEKCKGHWARDCPK--PRGPRGPRP 529

RESULT 10  
P3K3\_DICDI  
ID P3K3\_DICDI STANDARD; PRT; 1585 AA.  
AC P54675;  
DT 01-OCT-1996 (Rel. 34, Created)  
DT 01-OCT-1996 (Rel. 34, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE Phosphatidylinositol 3-kinase 3 (EC 2.7.1.137) (PI3-kinase)  
DE (PtdIns-3-kinase) (PI3K) (Fragment).  
GN PI3K OR PIK3.  
OS Dictyostellium discoideum (Slime mold).  
OC Eukaryota; Mycetozoa; Dictyosteliida; Dictyostelium.  
OX NCBI\_Taxid=44689;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=AX3;  
RX MEDLINE=96009592; PubMed=7565716;  
RA Zhou K., Takedawa K., Ehr S.D., Firtel R.A.;  
RT "A phosphatidylinositol (PI) kinase gene family in Dictyostelium  
discoideum: biological roles of putative mammalian p110 and yeast

RT Vps34p PI 3-kinase homologs during growth and development.\*;  
RL Mol. Cell. Biol. 15:5645-5656(1995).  
CC -1- CATALYTIC ACTIVITY: ATP + 1-phosphatidyl-1D-myo-inositol - ADP +  
CC 1-phosphatidyl-1D-myo-inositol 3-phosphate.  
CC -1- SIMILARITY: BELONGS TO THE PI3/PI4-KINASES FAMILY.  
CC -----  
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/  
CC or send an email to license@isb-sib.ch).  
CC -----  
CC EMBL; U23478; AAA85723.1; -  
DR Dictydb; D001101; PI3K.  
DR InterPro; IPR002420; PI3K\_C2.  
DR InterPro; IPR000341; PI3K\_ras\_bind.  
DR InterPro; IPR001263; PI3Ka.  
DR InterPro; IPR000403; PI3\_P14\_kinase.  
DR Pfam; PF00613; PI3Ka; 1.  
DR Pfam; PF00792; PI3K\_C2; 1.  
DR Pfam; PF00794; PI3K\_rbd; 1.  
DR Pfam; PF00454; PI3\_P14\_kinase; 1.  
DR SMART; SM00142; PI3K\_C2; 1.  
DR SMART; SM00144; PI3K\_rbd; 1.  
DR SMART; SM00145; PI3Ka; 1.  
DR SMART; SM00146; PI3K; 1.  
DR PROSITE; PS00915; PI3\_4\_KINASE\_1; 1.  
DR PROSITE; PS00916; PI3\_4\_KINASE\_2; 1.  
DR PROSITE; PS0290; PI3\_4\_KINASE\_3; 1.  
KW Transferase; Kinase; Multigene family; Repeat.  
FT NON\_TER 1 1  
FT DOMAIN 58 84 POLY-ASN.  
FT DOMAIN 98 110 POLY-ASN.  
FT DOMAIN 200 226 POLY-ASN.  
FT DOMAIN 239 254 POLY-ASN.  
FT DOMAIN 345 378 POLY-ASN.  
FT DOMAIN 383 390 POLY-ASN.  
FT DOMAIN 484 488 POLY-SER.  
FT DOMAIN 720 737 POLY-GLN.  
FT DOMAIN 1569 1575 POLY-ASN.  
FT DOMAIN 1221 1484 PI3K/PI4K.  
FT DOMAIN 1510 1539 5 X 5 AA APPROXIMATE REPEATS.  
FT REPEAT 1510 1514 1.  
FT REPEAT 1515 1519 2.  
FT REPEAT 1520 1524 3.  
FT REPEAT 1530 1534 4.  
FT REPEAT 1535 1539 5.  
FT DOMAIN 1547 1560 7 X 2 AA TANDEM REPEATS OF K-E.  
SQ SEQUENCE 1585 AA; 180421 MW; 4689B620D2484961 CRC64;

Query Match 5.7%; Score 121.5; DB 1; Length 1585;  
Best Local Similarity 21.0%; Pred. No. 2.7;  
Matches 71; Conservative 50; Mismatches 114; Indels 103; Gaps 17;

QY 51 PPTPPP-----NNNNNNKHTGHSACVPNMTRRRDELSEIINNLEKVMQSEEE 102  
DB 333 PSSSPPTQSDIFNENNNNNNNNNNNNN-----NNNNNNNNNNNNNNNEELINNNN 387  
QY 103 NNLLSQVQKLTENTTLREQVEP-----TPDEDDDDIELRG-----AAA 142  
DB 388 NNN--DENYKIETEESLKELEKELENEEREKILKERNEIDNKKHLSKGYFMRAC 445  
QY 143 AAAPPPPIEECEPDLP---EKFDGNDPMLAPMAQCQIFMEKSTDFSVDRVRCVTS 199  
DB 446 NASNDDGLEE---EDIPLOQDEHWETNVILLPCRHHVKVPCSSSS---SIDS- 492  
QY 200 MMTGRAARWASAKLERSHYLMHNYPAFMMEM--KHVEDPQRREVAKRRLRROGMSV 257  
DB 493 ----RLAWASGKMQHNLNLEKDEKFFTLRCMKNKVYFDQD-----TPIGLH 535

```
QY 258 IDYSNAFOMIAQDLWNEPA-----LIDYHEGLSDHIQEE-----LSHLEVAKSLS 304
Db 536 IQY-----NLANNPTOKPTNIKLELV-----LEDELCKRLVDLOSLEINNGRP 580
QY 305 AL-----IGOCIHIERRLARAAARPRS--PRALVLP 337
Db 581 SIWKSHIDVDFNSRKLRELAKRQSNVPAARLTPY 618

RESULT 11
Y182_HUMAN
ID Y182_HUMAN STANDARD; PRT; 1157 AA.
AC Q14687;
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical protein KIAA0182 (Fragment).
GN KIAA0182.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Bone marrow;
RX MEDLINE=96281124; PubMed=8724849;
RA Nagase T., Seki N., Ishikawa K.-I., Tanaka A., Nomura N.;
RT "Prediction of the coding sequences of unidentified human genes. V.
RT The coding sequences of 40 new genes (KIAA0161-KIAA0200) deduced by
RT analysis of cDNA clones from human cell line KG-1.";
RL DNA Res. 3:17-24(1996).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; D80004; BA01499.1; -
KW Hypothetical protein.
FT NON_TER 1
FT DOMAIN 591 596 POLY-PRO.
FT DOMAIN 685 688 POLY-ARG.
FT DOMAIN 1042 1047 POLY-GLU.
FT SEQUENCE 1157 AA; 130323 MW; B0DADA5A6FAB5FB CRC64;
SQ
Query Match 5.6%; Score 120.5; DB 1; Length 1157;
Best Local Similarity 21.9%; Pred. No. 2.1;
Matches 95; Conservative 52; Mismatches 124; Indels 163; Gaps 25;
QY 23 LPHRSEATAGSPPTPTVT-----LGDPCPP-----PPPPP-----PNNNNNN 63
Db 354 LHGLRGHATEERGKPSQELTPTRAEKLDASLQAPKPVQHPHPVTPHTTVPVSLSNHG 413
QY 64 SKHTGKSCACVNNWTRRDELSEINNLRKVMQSENNNLQSQVKLTFTNTLREQ 123
Db 414 IFSLPSSAATALLIORTNE-----EEKLARQRLRQEKEDR--QSQVSEFRQO--VLEQH 466
QY 124 VE-----PTPEDDDIELRGAAAPPPPIEECPEDLPKFDGNDPDLAP-----FNAQ 175
Db 467 LDMGRPPVPAEAHRPE-----STRPGNRHPEGRDPDQHPGPPPLISPKPQLHAAP 521
QY 176 QOIFMEKSTDFSVDRVRVCFVTSMVTGRAARWASAKLERSHYLMHNPAP----- 226
Db 522 TALWNPVSLMD-----NTLETRA-----ESHSL-HSHPAAPPEPSRQAQV 560
QY 227 NMENKHFVQPRREVAKRKRRLRQMGSVIDYSNFAQMTAQDLWNEPALIDQY--- 282
Db 561 PLVKVERVF-CPEKAEGRK-----REPAPLDKYQPP 592
```

```
QY 283 -----HEGLSDHIQEEELSH-----LEVAKSLSALIGQCIIHIERRLARAAAR--- 324
Db 593 PPPPPGGS-----LEHQPLPGPGPFLAELEKSTQTILGQ-----QRASLPQAATFGE 641
QY 325 -----KPRSP-----PRALVLPFIASHHQVDFT-----EPVGGARMRLTQE---EKERRKL 368
Db 642 LSGPLKPGSPYRPPVPRA-----PDPAYIDFLOQRRRLVSKLDLEERRRREA 690
QY 369 NLCLYCGTGGHYAD 382
Db 691 Q-----EKGIYYID 698

RESULT 12
AF6_HUMAN
ID AF6_HUMAN STANDARD; PRT; 1816 AA.
AC P55196; O75087; O75088; O75089; Q9NU92;
DT 01-OCT-1996 (Rel. 34, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DE AF-6 protein.
GN MLT4 OR AF6.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A. (ISOFORM 1).
RX MEDLINE=94061833; PubMed=8242616;
RA Prasad R., Gu Y., Alder H., Nakamura T., Canaan O., Saito H.,
RA Huebner K., Gale R.P., Nowell P.C., Kuriyama K., Miyazaki Y.,
RA Croce C.M., Canaan O.
RT "Cloning of the ALL-1 fusion partner, the AF-6 gene, involved in
RT acute myeloid leukemias with the t(6;11) chromosome translocation.";
RL Cancer Res. 53:5624-5628(1993).
RN [2]
RP SEQUENCE FROM N.A., AND ALTERNATIVE SPLICING.
RC TISSUE=Fetal brain;
RX MEDLINE=98344142; PubMed=9679199;
RA Saito S., Matsushima M., Shirahama S., Minaguchi T., Kanamori Y.,
RA Minami M., Nakamura Y.;
RT "Complete genomic structure, DNA polymorphisms, and alternative
RT splicing of the human AF-6 gene.";
RL DNA Res. 5:115-120(1998).
RN [3]
RP SEQUENCE OF 337-1816 FROM N.A. (ISOFORM 2).
RA Williams S.;
RL Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: MAY ACT AS AN INTRACELLULAR SIGNALING COMPONENT
CC CONTROLLED BY RAS SIGNALING PATHWAYS.
CC -!- SUBUNIT: BINDS DIRECTLY TO ZO-1 AND OCCLUDIN.
CC -!- SUBCELLULAR LOCATION: Cytoplasmic (Probable).
CC -!- ALTERNATIVE PRODUCTS: 3 ISOFORMS; 1, 2 (SHOWN HERE) AND 3; ARE
CC PRODUCED BY ALTERNATIVE SPLICING.
CC -!- DISEASE: INVOLVED IN ACUTE LEUKEMIAS BY A CHROMOSOMAL
CC TRANSLOCATION T(6;11)(Q27;Q23) THAT INVOLVES MLT4 AND MLL/HRX.
CC THE RESULT IS A ROGUE ACTIVATOR PROTEIN.
CC -!- SIMILARITY: CONTAINS 1 DILUTE DOMAIN.
CC -!- SIMILARITY: CONTAINS 1 FHA DOMAIN.
CC -!- SIMILARITY: CONTAINS 1 PDZ/DHR DOMAIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AB011399; BAA32484.1; -
DR EMBL; AB011399; BAA32483.1; -
DR EMBL; AB011399; BAA32485.1; -
DR EMBL; U02478; AAC50059.1; -
```

```

DR EMBL; AL049698; CAB76850.1; -.
DR HSSP; Q12923; 3PDZ.
DR MIM; 159559; -.
DR InterPro; IPR002710; DIL.
DR InterPro; IPR000253; FHA_domain.
DR InterPro; IPR001478; PDZ.
DR InterPro; IPR000159; RA.
DR Pfam; PF01843; DIL; 1.
DR Pfam; PF00498; FHA; 1.
DR Pfam; PF00595; PDZ; 1.
DR Pfam; PF00788; RA; 2.
DR ProDom; PD003376; DIL; 1.
DR SMART; SM00240; FHA; 1.
DR SMART; SM00228; PDZ; 1.
DR SMART; SM00314; RA; 2.
DR PROSITE; PS50106; PDZ; 1.
KW Chromosomal translocation; Proto-oncogene; Alternative splicing.
FT DOMAIN 36 206 RAS-INTERACTING.
FT DOMAIN 425 491 FHA.
FT DOMAIN 804 910 DILUTE.
FT DOMAIN 991 1077 PDZ.
FT DOMAIN 162 174 GLU/LYS-RICH.
FT DOMAIN 1349 1356 POLY-PRO.
FT DOMAIN 1371 1376 POLY-PRO.
FT DOMAIN 1561 1571 ASP/GLU-RICH (ACIDIC).
FT SITE 26 ML FUSION POINT (IN AN ACUTE MYELOID LEUKEMIA PATIENT).
FT VARSPLIC 1588 1611 LODEERRROOOLEEMKREAEADRA -> VKGGVWLCPSPV
FT VARSPLIC 1612 1816 MISSING (IN ISOFORM 1).
FT VARSPLIC 1666 1743 SLRPLPRDYEPSPAPGAPPPPPQNRNASYLKTVQLSPD
LFTAKFVAYNEEEEDCSLAGQPKYSTRKSHGDL ->
PNSYPGSTGAAYGAHDACDRKSKSDADPGSSGAPE
NTEKERQLFSQGDVSNKVKASRKLTELENLTK (IN
ISOFORM 3).
FT VARSPLIC 1744 1816 MISSING (IN ISOFORM 3).
FT CONFLICT 373 373 G -> V (IN REF. 1).
FT CONFLICT 391 391 P -> PGRNHFAYNHYTYE (IN REF. 3).
FT CONFLICT 744 744 D -> DSHFD (IN REF. 3).
FT CONFLICT 1031 1031 D -> DV (IN REF. 1).
FT CONFLICT 1408 1408 R -> P (IN REF. 1).
FT SEQUENCE 1816 AA; 205604 MW; EB1FE7F04879CE8F CRC64;

Query Match 5.6%; Score 120.5; DB 1; Length 1816;
Best Local Similarity 19.7%; Pred. No. 3.6;
Matches 70; Conservative 46; Mismatches 159; Indels 81; Gaps 11;

QY 31 TAGRSPTTVTLGP---DCPPPPPPPPNNNN-----NNSKHTGHSACV 74
Db 1328 TPAAIPTAVVAVSQPIRTDLPPPPPPVHYAGDFGMSMDLPLPPPPSANQIGLPSAQV 1387
QY 75 PNMTERRRDELSSEINLNREKVMKQSENNNLSQVOKLTTEENTTLREQVEPTPEDDDD 134
Db 1388 AAARERKREHORWYE--KEKARLEERERKRREQERKLGQMR-----QSLNPAFFSPLTA 1442
QY 135 IELRGAAAAAAPP---TEEECPDLPEKFGNPDMLAPPMAQCOIFMEKSTRDFSVD 191
Db 1443 QQMKPEKPTLQPOETVIRELPQQQPTIE-----RRDLQYITVSK 1485
QY 192 VRVCFVTSMTGAARWAKSALERSHYLMHNYPAFMWEKHWPEDPQREVAKRKRRLR 251
Db 1486 EELSGSDLSPPDPKWRDRAKLEKQKQ--MHIVDMLSKETQLSQKSDRKAESDRKLKM 1544
QY 252 QGMSGVIDYSNAFQMTAQLDWNNEPALIDQYHGLSDHTQEEHLSHVAKSLSALLGOCI 311
Db 1545 -----LEWQFKRLQESKQKDEDEDEDDDDVD-----TMLIMQRL 1580
QY 312 HIERRLAAAAARKPRSPRALVPHIASHHQVDPTPEVGARMRLTQEEKERRK 367
Db 1581 EAERRARLQDEERR-----RQOOLEEMKREAE-----RARQEERRRQE 1621

```

```

RESULT 13
CHDM DROME
ID CHDM_DROME STANDARD; PRT; 1982 AA.
AC Q97159; Q9VW50.
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Chromodomain helicase-DNA-binding protein Mi-2 homolog (dmi-2).
GN MI-2 OR CG8103.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A., FUNCTION, AND MUTAGENESIS OF GLY-737.
RX MEDLINE=99055400; PubMed=9836641;
RA Kehle J., Beuchle D., Treuhelt S., Christen B., Kennison J.A.,
RA Bienz M., Muller J.;
RT "dmi-2, a hunchback-interacting protein that functions in Polycomb
repression.";
RL Science 282:1897-1900(1998).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=BERKELEY;
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA Sutton G.G., Wortman J.R., Richards S., Ashburner M., Henderson S.N.,
RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abril J.F., Aghayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Bereman B.P., Bhandari D., Bolshakov S.,
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brotter P.,
RA Burkis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Fostler K.C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Minkov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reiner K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svirskaas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wassarman D.A., Weinstein G.M., Weissenbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zhong X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of Drosophila melanogaster.";
RL Science 287:2185-2195(2000).
CC -!- FUNCTION: VITAL ROLE IN DEVELOPMENT. PROTEIN BINDS TO A PORTION OF
HUNCHBACK (HB) PROTEIN THAT IS CRITICAL FOR REPRESSION OF BITHORAX
COMPLEX (BXC) GENES. MAY ALSO FUNCTION IN POLYCOMB GROUP (PCG)
REPRESSION OF HOX GENES.
CC -!- SUBCELLULAR LOCATION: Nuclear.
CC -!- SIMILARITY: BELONGS TO THE SNF2/RAD54 HELICASE FAMILY.
CC -!- SIMILARITY: CONTAINS 2 'CHROMO' DOMAINS.
CC -!- SIMILARITY: CONTAINS 2 PHD-TYPE ZINC FINGERS.

```

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC -----  
 DR EMBL; AF119716; AAD17276.1; -  
 DR EMBL; AE003515; AAF49099.1; ALT\_SEQ.  
 DR FlyBase: FBgn0013591; Mi-2.  
 DR InterPro: IPR000953; Chromo.  
 DR InterPro: IPR001410; DEAD.  
 DR InterPro: IPR002464; DEAH\_ATP\_helicase.  
 DR InterPro: IPR001650; Helicase\_C.  
 DR InterPro: IPR001965; PHD.  
 DR InterPro: IPR000330; SNF2\_N.  
 DR InterPro: IPR001841; Zn\_finger.  
 DR Pfam: PF00385; chromo; 1.  
 DR Pfam: PF00271; helicase\_C; 1.  
 DR Pfam: PF00628; PHD; 2.  
 DR Pfam: PF00176; SNF2\_N; 1.  
 DR SMART; SM00298; CHROMO; 2.  
 DR SMART; SM00487; DEXDC; 1.  
 DR SMART; SM00490; HELICC; 1.  
 DR SMART; SM00249; PHD; 2.  
 DR SMART; SM00184; RING; 2.  
 DR PROSITE; PS00598; CHROMO\_1; FALSE\_NEG.  
 DR PROSITE; PS00690; DEAH\_ATP\_HELICASE; 1.  
 DR PROSITE; PS00013; CHROMO\_2; 2.  
 KW DNA-binding; ATP-binding; Helicase; Nuclear protein; Repeat;  
 KW Transcription regulation; Repressor; Zinc-finger.  
 FT ZN\_FING 380 421 PHD-TYPE.  
 FT ZN\_FING 440 481 PHD-TYPE.  
 FT DOMAIN 488 566 CHROMO 1.  
 FT DOMAIN 512 573 CHROMO 2.  
 FT NP\_BIND 755 762 ATP (POTENTIAL).  
 FT SITE 875 878 DEAH BOX.  
 FT DOMAIN 13 16 POLY-GLU.  
 FT DOMAIN 70 76 POLY-LYS.  
 FT DOMAIN 239 248 POLY-GLU.  
 FT DOMAIN 1279 1287 POLY-GLU.  
 FT DOMAIN 1672 1677 POLY-ASP.  
 FT MUTAGEN 737 737 G->D: IN ALLELE MI-2-5; LARVAL LETHAL.  
 FT CONFLICT 101 101 G -> A (IN REF. 1).  
 SQ SEQUENCE 1982 AA; 224199 MW; ED8E256D1AD0AC2F CRC64;

Query Match 5.6%; Score 120.5; DB 1; Length 1982;  
 Best Local Similarity 18.8%; Pred. NO. 4;  
 Matches 80; Conservative 63; Mismatches 152; Indels 131; Gaps 16;  
 QY 27 RSEATAGRSP-----TPTVTLGDCPPPPPPPPNNNNNNKSHGKHSACVPM 77  
 DB 1529 KDVAALAEAPPNGVNDKATTSVTSATSAAPAPASEKGDCKD----- 1577  
 QY 78 TERRDELSEETNNLRKVMKQSENN-----NLQSOVKILT 114  
 DB 1578 SEKEDKTSAEKSEVKQE--QEAEDKPKGVQENPVEEAGDTKPSDAEVKTEVAKTE 1635  
 QY 115 ENTTLRQVEPTPEDEDDIELRGAAAAAAPPPIEECECPEDLPKFDGNPDMLAPFMA 174  
 DB 1636 KPEETKDPVEKPEETKEKEKVDKPKPIPTVTVIDDDDDVMIVKEDGE----- 1686  
 QY 175 QCOIFMEKSTRDFSVDYRVVCFVTSMITGRAARWASAKLE--RSHVLMHNYPAFWMEMKH 232  
 DB 1687 -----LEKPSASSPKDKKAAARAAATSAATGATGKGAEDSLVUKRFMFNIADGGTTLHT 1741  
 QY 233 VFEDPQRREVARKKI-----LRQGMGSVIDSN--AFQMIAQDLDPWNEPA 277  
 DB 1742 LWLNKEKAAPVGREVEIWHRRHRYWLLAGIVTHGYGRQDQINDIRFAII-----NEPF 1795  
 QY 278 LIDQVHEGLSDHIQELSHLEVAKLSA-----LIGQCIIHERRLRAAAARKPRSPRAL 333

DB 1796 KMD-VCKG-----NFEIKNFKLARRFKLLEQALVIEQLRRAAYLNLAQ----- 1839  
 QY 334 VLPFIASHHQVDPTEPVGGARMRLTQEEKERRKLNLCYCGTGGHYADCNPAKASKSSP 393  
 DB 1840 -----DPSHPAMSLNARFAVE-----CLAESH-----QHLKSLEL 1870  
 QY 394 AGNSPA 399  
 DB 1871 AGNKPA 1876  
 RESULT 14  
 ID HGL2\_ARATH STANDARD; PRT; 745 AA.  
 AC P46607; Q39018;  
 DT 01-NOV-1995 (Rel. 32, Created)  
 DT 01-NOV-1997 (Rel. 35, Last sequence update)  
 DE 01-MAR-2002 (Rel. 41, Last annotation update)  
 DE Homeobox protein GLABRA2 (Homeobox-leucine zipper protein ATHB-10)  
 DE (HD-ZIP protein ATHB-10).  
 GN GL2 OR AT1G79840 OR F19K16.20.  
 OS Arabidopsis thaliana (Mouse-ear cress).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;  
 OC eurosids II; Brassicales; Brassicaceae; Arabidopsids.  
 OX NCBI\_TaxID=3702;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=CV. WASSILEWSKIJA; TISSUE=Seedling;  
 RX MEDLINE=95011550; PubMed=7926739;  
 RA Rerle W.G., Feldmann K.A., Marks M.D.;  
 RT "The GLABRA2 gene encodes a homeo domain protein required for normal  
 RL trichome development in Arabidopsis."; Genes Dev. 8:1388-1399(1994).  
 RN [2]  
 RP REVISIONS.  
 RA Marks M.D.;  
 RL Submitted (NOV-1996) to the EMBL/GenBank/DBJ databases.  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=CV. COLUMBIA;  
 RX MEDLINE=96407838; PubMed=8811855;  
 RA di Cristina M., Sessa G., Dolan L., Linstead P., Baima S., Ruberti I.,  
 RT Morelli G.;  
 RL "The Arabidopsis Athb-10 (GLABRA2) is an HD-zip protein required for  
 regulation of root hair development."; Plant J. 10:393-402(1996).  
 RN [4]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=CV. COLUMBIA;  
 RX MEDLINE=21016719; PubMed=11130712;  
 RA Theologis A., Ecker J.R., Palm C.J., Federspiel N.A., Kaul S.,  
 White O., Alonso J., Altafi H., Araujo R., Bowman C.L., Brooks S.Y.,  
 Chugh E., Chen H., Chao Q., Chen K., Cheuk R.F., Chin C.W.,  
 Chung M.K., Conn L., Conway A.B., Conway A.R., Creasy T.H., Dewar K.,  
 Dunn P., Etgu P., Feldblyum T.V., Feng J.-D., Fong B., Fujii C.Y.,  
 Gill J.E., Goldsmith A.D., Haas B., Hansen N.F., Hughes B., Huizar L.,  
 Hunter J.L., Jenkins J., Johnson-Hopson C., Khan S., Khaykin E.,  
 Kim C.-J., Koo H.-B., Kremenetskaia I., Kurtz D.B., Kwan A., Lam B.,  
 Langin-Hooper S., Lee A., Lee J.M., Lenz C.A., Li J.H., Li Y.-P.,  
 Lin X., Liu S.X., Liu Z.A., Lueros J.S., Maiti R., Marziani A.,  
 Militscher J., Miranda M., Nguyen M., Nierman W.C., Osborne B.I.,  
 Pai G., Peterson J., Pham P.K., Rizzo M., Rooney T., Rowley D.,  
 Sakano H., Salzberg S.L., Schwartz J.R., Shinn P., Southwick A.M.,  
 Sun H., Tallon L.J., Tambunga G., Toriumi M.J., Town C.D.,  
 Uterback T., Van Aken S., Vaysberg M., Vysotskaya V.S., Walker M.,  
 Wu D., Yu G., Fraser C.M., Venter J.C., Davis R.W.;  
 RT "Sequence and analysis of chromosome 1 of the plant Arabidopsis  
 thaliana";  
 RL Nature 408:816-820(2000).  
 CC -!- FUNCTION: REQUIRED FOR CORRECT MORPHOLOGICAL DEVELOPMENT AND  
 CC MATURATION OF TRICHOMES AS WELL AS FOR NORMAL DEVELOPMENT OF SEED







GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 13, 2002, 15:09:24 ; Search time 31.91 Seconds  
(without alignments)  
2173.957 Million cell updates/sec

Title: US-09-631-863A-2  
Perfect score: 2139  
Sequence: 1 MRNKRVLTKTKRRSGRGQD.....DNCPAKASKSSPAGNSPAPL 401

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5  
Searched: 562222 seqs, 172994929 residues  
Total number of hits satisfying chosen parameters: 562222

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : SPTREMBL19:\*

- 1: sp\_archaea:\*
- 2: sp\_bacteria:\*
- 3: sp\_fungi:\*
- 4: sp\_human:\*
- 5: sp\_invertebrate:\*
- 6: sp\_mammal:\*
- 7: sp\_mhc:\*
- 8: sp\_organelle:\*
- 9: sp\_phase:\*
- 10: sp\_plant:\*
- 11: sp\_rodent:\*
- 12: sp\_virus:\*
- 13: sp\_vertebrate:\*
- 14: sp\_unclassified:\*
- 15: sp\_virus:\*
- 16: sp\_bacteriap:\*
- 17: sp\_archeap:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	1802	84.2	342	4 Q9UPV1	Q9upv1 homo sapien
2	1703	79.6	325	4 Q96A68	Q96a68 homo sapien
3	943	44.1	231	11 Q9EQ11	Q9eq11 mus musculu
4	373	17.4	704	13 Q98SV9	Q98sv9 fugu rubrip
5	364.5	17.0	371	13 Q93283	Q93283 fugu rubrip
6	279	13.0	1252	3 Q96V43	Q96v43 aspergillus
7	240.5	11.2	853	3 Q00833	Q00833 fusarium ox
8	229.5	10.7	837	3 Q9HFY8	Q9hfy8 collettotric
9	223.5	10.4	400	6 Q95KK1	Q95kk1 macaca fasc
10	192.5	9.0	1418	4 Q9P2M8	Q9p2m8 homo sapien
11	189	8.8	1810	10 Q94115	Q94i15 oryza sativ
12	185	8.6	1500	10 Q94DA0	Q94da0 oryza sativ
13	185	8.6	1862	10 Q94104	Q94i04 oryza sativ
14	184.5	8.6	1781	10 Q9AYC0	Q9ayc0 oryza sativ
15	182.5	8.5	1473	10 Q94H22	Q94h22 oryza sativ
16	180.5	8.4	959	10 Q94108	Q94i08 oryza sativ

17	179.5	8.4	1714	10 Q94HT9	Q94ht9 oryza sativ
18	179.5	8.4	1777	10 Q94HP9	Q94hp9 oryza sativ
19	179	8.4	2017	10 Q9AYB6	Q9ayb6 oryza sativ
20	178.5	8.3	1491	10 Q9XE26	Q9xe26 oryza sativ
21	172.5	8.1	1773	10 Q94HV6	Q94hv6 oryza sativ
22	170.5	8.0	1506	10 Q94DR3	Q94dr3 oryza sativ
23	167	7.8	1524	10 Q9LWJ0	Q9lwj0 oryza sativ
24	166.5	7.8	1475	10 Q94H45	Q94h45 oryza sativ
25	164.5	7.7	1571	10 Q94HN3	Q94hn3 oryza sativ
26	163	7.6	1521	10 Q9XEM5	Q9xem5 oryza sativ
27	163	7.6	2157	10 Q9AYB5	Q9ayb5 oryza sativ
28	162.5	7.6	406	3 Q9UVD9	Q9uvd9 alternaria
29	161.5	7.6	2162	10 Q9AYC2	Q9ayc2 oryza sativ
30	161	7.5	1220	10 Q9XHY2	Q9xhy2 oryza sativ
31	161	7.5	1521	10 Q94DL1	Q94dl1 oryza sativ
32	159.5	7.5	639	3 Q9UVC2	Q9uvc2 cladosporiu
33	159.5	7.5	963	10 Q9LWX0	Q9lwx0 oryza sativ
34	157	7.3	1524	10 Q93VN0	Q93vn0 oryza sativ
35	156	7.3	1461	10 Q94H43	Q94h43 oryza sativ
36	151	7.1	372	3 Q9Y8H2	Q9y8h2 tricholoma
37	148	6.9	1494	5 Q94885	Q94885 drosophila
38	146	6.8	1524	10 Q943A1	Q943a1 oryza sativ
39	142.5	6.7	303	4 Q9H7A0	Q9h7a0 homo sapien
40	139	6.5	1587	10 Q9SLQ0	Q9slq0 oryza sativ
41	138.5	6.5	2075	13 Q90WA4	Q90wa4 fugu rubrip
42	138	6.5	675	11 Q91YM9	Q91ym9 mus musculu
43	138	6.5	1123	11 Q9DBD5	Q9dbd5 mus musculu
44	137	6.4	498	10 Q94BM0	Q94bm0 hordeum vul
45	135	6.3	100	11 Q9DCZ3	Q9dcz3 mus musculu

ALIGNMENTS

RESULT 1

Q9UPV1 PRELIMINARY; PRT; 342 AA.

AC Q9UPV1; DT 01-NAY-2000 (TREMBlrel. 13, Created)

DT 01-NAY-2000 (TREMBlrel. 13, Last sequence update)

DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)

DE KIAA1051 PROTEIN (FRAGMENT).

GN KIAA1051.

OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

OX NCBI\_TaxID=9606;

RN [1]

RP SEQUENCE FROM N.A.

RC TISSUE=BRAIN;

RX MEDLINE=9397452; PubMed=10470851;

RA Kikuno K., Nagase T., Ishikawa K., Hirotsawa M., Miyajima N., Tanaka A., Kotani H., Nomura N., Ohara O.;

RT "Prediction of the coding sequences of unidentified human genes. XIV. The complete sequences of 100 new cDNA clones from brain which code for large proteins in vitro.";

RL DNA Res 6:197-205(1999).

DR EMBL; AB028974; BAA83003.1; -.

DR InterPro; IPR001878; Znf\_CCHC.

DR Pfam; PF00098; zf-CCHC; 1.

DR SMART; SM00343; Znf\_C2HC; 1.

KW Zinc-finger.

FT NON\_TER

SQ SEQUENCE 342 AA; 38769 MW; DC14C265B92541EB CRC64;

Query Match 84.2%; Score 1802; DB 4; Length 342;  
Best Local Similarity 100.0%; Pred. No. 7.8e-131;  
Matches 342; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 60 NNNNSKHTGKHSACVPNMTERRRDELSEENLNLRKVMKQSENNNLQSOVKLTTEENTT 119  
|||||

Db 1 NNNNSKHTGKHSACVPNMTERRRDELSEENLNLRKVMKQSENNNLQSOVKLTTEENTT 60  
|||||

Qy 120 LREQVEPTPEDEDDIEDLARGAAAAAPPPPIEECEPEDLPKFGDGNPDMLAPFMAQCOIF 179  
 |||  
 Db 61 LREQVEPTPEDEDDIEDLARGAAAAAPPPPIEECEPEDLPKFGDGNPDMLAPFMAQCOIF 120  
 |||  
 Qy 180 MEKSTRDFSVDRVRVCFVTSMTGRAARWASAKLERSHYLMHNPAPFMEMKHVFPEDPQR 239  
 |||  
 Db 121 MEKSTRDFSVDRVRVCFVTSMTGRAARWASAKLERSHYLMHNPAPFMEMKHVFPEDPQR 180  
 |||  
 Qy 240 REVARKIRRLROGMSVIDYSNAPQMTAQLDWNNEPALIDQYHEGLSDHIOEELSHLEV 299  
 |||  
 Db 181 REVARKIRRLROGMSVIDYSNAPQMTAQLDWNNEPALIDQYHEGLSDHIOEELSHLEV 240  
 |||  
 Qy 300 AKSLGALIGOCIHIERRLARAAAAARKPRSPRALVPLPHIASHHQVDPTEPVGGARMRLTQ 359  
 |||  
 Db 241 AKSLGALIGOCIHIERRLARAAAAARKPRSPRALVPLPHIASHHQVDPTEPVGGARMRLTQ 300  
 |||  
 Qy 360 BEKERRRLNLCCLXCGTGGHYADNCPAKASKSSPAGNSPAPL 401  
 |||  
 Db 301 BEKERRRLNLCCLXCGTGGHYADNCPAKASKSSPAGNSPAPL 342  
 |||

## RESULT 2

Q96A68 ID Q96A68 PRELIMINARY; PRT; 325 AA.  
 AC Q96A68;  
 DT 01-DEC-2001 (Tremblrel. 19, Created)  
 DT 01-DEC-2001 (Tremblrel. 19, Last sequence update)  
 DE ME3 LIKE 1 (Paternally expressed gene 10 ORF1).  
 GN ME3L1 OR PEG10.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE-TESTIS;  
 RA Nakamura Y., Furukawa Y.;  
 RL Submitted (SEP-2000) to the EMBL/GenBank/DBJ databases.  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE-TESTIS;  
 RA Satoh S., Furukawa Y.;  
 RL "Isolation of ME3 like gene 1.";  
 RT Submitted (SEP-2001) to the EMBL/GenBank/DBJ databases.  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=21218923; PubMed=11318613;  
 RA Ono R., Kobayashi S., Wagatsuma H., Aisaka K., Kohda T.,  
 RA Kaneko-Ishino T., Ishino F.;  
 RT "A Retrotransposon-Derived Gene, PEG10, Is a Novel Imprinted Gene  
 RT Located on Human Chromosome 7q21.";  
 RL Genomics 73:232-237(2001).  
 DR EMBL; AB049150; BAB68387.1; -;  
 DR EMBL; AB049834; BAB43951.1; -;  
 SQ SEQUENCE 325 AA; 36965 MW; 118E4CAF97F2A76 CRC64;

Query Match 79.68; Score 1703; DB 4; Length 325;  
 Best Local Similarity 100.0%; Pred. No. 3e-123;  
 Matches 325; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 77 MTERRRDSEINLRKVMKQSENNNLSQVQKLTEENTTLREOVEPTPEDEDDIE 136  
 |||  
 Db 1 MTERRRDSEINLRKVMKQSENNNLSQVQKLTEENTTLREOVEPTPEDEDDIE 60  
 |||  
 Qy 137 LRGAAAAAPPPPIEECEPEDLPKFGDGNPDMLAPFMAQCOIFMEKSTRDFSVDRVRVCF 196  
 |||  
 Db 61 LRGAAAAAPPPPIEECEPEDLPKFGDGNPDMLAPFMAQCOIFMEKSTRDFSVDRVRVCF 120  
 |||  
 Qy 197 VTSMTGRAARWASAKLERSHYLMHNPAPFMEMKHVFPEDPQRREVAKRKIRRLROGMS 256  
 |||

Db 121 VTSMTGRAARWASAKLERSHYLMHNPAPFMEMKHVFPEDPQRREVAKRKIRRLROGMS 180  
 |||  
 Qy 257 VIDYSNAPQMTAQLDWNNEPALIDQYHEGLSDHIOEELSHLEVAKSLGALIGOCIHIERR 316  
 |||  
 Db 181 VIDYSNAPQMTAQLDWNNEPALIDQYHEGLSDHIOEELSHLEVAKSLGALIGOCIHIERR 240  
 |||  
 Qy 317 LARAAAAARKPRSPRALVPLPHIASHHQVDPTEPVGGARMRLTQBEKERRRLNLCCLYCGT 376  
 |||  
 Db 241 LARAAAAARKPRSPRALVPLPHIASHHQVDPTEPVGGARMRLTQBEKERRRLNLCCLYCGT 300  
 |||  
 Qy 377 GGHYADNCPAKASKSSPAGNSPAPL 401  
 |||  
 Db 301 GGHYADNCPAKASKSSPAGNSPAPL 325  
 |||

## RESULT 3

Q9EQ11 ID Q9EQ11 PRELIMINARY; PRT; 231 AA.  
 AC Q9EQ11;  
 DT 01-MAR-2001 (Tremblrel. 16, Created)  
 DT 01-DEC-2001 (Tremblrel. 19, Last sequence update)  
 DE MYELIN EXPRESSION FACTOR-3-LIKE PROTEIN (FRAGMENT).  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OX NCBI\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=C57BL/6;  
 RX MEDLINE=21105984; PubMed=11158386;  
 RA Wolff J.-N., Koerting C., Scharltl M.;  
 RT "Ty3/Gypsy retrotransposon fossils in mammalian genomes: did they  
 RT evolve into new cellular functions?";  
 RL Mol. Biol. Evol. 18:266-270(2001).  
 DR EMBL; AF302691; AAC39979.1; -;  
 DR InterPro; IPR001878; Znf\_CCHC.  
 DR Pfam; PF00098; zf\_CCHC; 1.  
 DR SMART; SM00343; Znf\_C2HC; 1.  
 FT NON\_FINGER. 1  
 FT NON\_TER. 1  
 SQ SEQUENCE 231 AA; 26171 MW; DEA82A2E624F3974 CRC64;

Query Match 44.1%; Score 943; DB 11; Length 231;  
 Best Local Similarity 77.0%; Pred. No. 5.9e-65;  
 Matches 181; Conservative 20; Mismatches 30; Indels 4; Gaps 2;

Qy 167 DMLAPFMAQCOIFMEKSTRDFSVDRVRVCFVTSMTGRAARWASAKLERSHYLMHNPAPF 226  
 |||  
 Db 1 DMLGPFMYQCQLEFMEKSTRDFSVDRIRVCFVTSMLIGRAARWATAKLRCTYLMHNTAF 60  
 |||  
 Qy 227 MMEMKHVFPEDPQRREVAKRKIRRLROGMSVIDYSNAPQMTAQLDWNNEPALIDQYHEGL 286  
 |||  
 Db 61 MMELKHVFPEDPQRREAAKRIIRLRQGGPVVDYSNAPQMTAQLDWTPEALMDQFQGL 120  
 |||  
 Qy 287 SDHIOEELSHLEVAKSLGALIGOCIHIERRLARAAAAARKPRSPRALVPLPHIASHHQVDP 346  
 |||  
 Db 121 NPDIRAELSROEAPKTLAALITACIHIERRLARAAAA-KPDSPRALVMP---PNSQIDP 176  
 |||  
 Qy 347 TEPVGGARMRLTQBEKERRRLNLCCLYCGTGGHYADNCPAKASKSSPAGNSPAPL 401  
 |||  
 Db 177 TEPVGGARMRLSKEEKERRRRKMNLCYCGNGCHEADTCAKASKNSPPGNSPAPL 231  
 |||

## RESULT 4

Q98SV9 ID Q98SV9 PRELIMINARY; PRT; 704 AA.  
 AC Q98SV9;  
 DT 01-JUN-2001 (Tremblrel. 17, Created)  
 DT 01-JUN-2001 (Tremblrel. 17, Last sequence update)  
 DT 01-DEC-2001 (Tremblrel. 19, Last annotation update)  
 DE GAG-PROTEASE.



```
RL Appl. Environ. Microbiol. 61:1068-1072(1995).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=NRRL 1997; TRANSPOSON=AFRTL-1;
RA Okubara P.A., Tibbot B., McAlpin C.E., Hua S.-S.T.;
RT "AfRL-1, a retrotransposon-like element in the aflatoxin-producing
RT fungus Aspergillus flavus.";
RL Submitted (MAR-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF362957; AAL26311.1; -.
FT NON_TER 1252 1252
FT SEQUENCE 1252 AA; 144752 MW; D8FDC5DD8D8E6C4E8 CRC64;
SQ SEQUENCE 853 AA; 99422 MW; 1AEDD18CFBBA5B06 CRC64;

Query Match 11.2%; Score 240.5; DB 3; Length 853;
Best Local Similarity 21.2%; Pred. No. 3.3e-10;
Matches 85; Conservative 65; Mismatches 157; Indels 93; Gaps 14;

QY 33 GRSPPTPTVTGLGPCPPPPPPPPNNNNNNNNKHTGHKSACVPNMTERRDEL---SEEI 89
DB 19 GHQPAAP-----ANPAPVNRPTDQMDADDSSOSSD--DSEVERLREOLGNVINEM 69
QY 90 NNIREKV-----MKOSENNNLQSOVKLTEENTTLR---EQVEPTPEDEDDIELRG 139
DB 70 NEMRQMLEEFTALHQOQNSNNNTQEQMYNLASAANNRDPGEVLKPS-----118
QY 140 AAAAAAPPIIEECPELDEKFDGNDMLAPFMAOQIEKSTDFSDVRVRCVFTS 199
DB 119 -----PEYFDGTPSKLPTFLTQSRAFITYYPNQFRNDSAKVMTMAG 159
QY 200 MMTGRAARWASAKLERSHYLMHNY---PAFMME-----MKHVFEDPQ 238
DB 160 RLITAAQWQP-----IMNDYMTNPYKLPQRTALLFGENGRHEMERALKMAFGTID 212
QY 239 RREVAKKIRRLROGMSGVIDYSNAFOMIAODLDWNEPALIDQYHEGLSDHIQELSLE 298
DB 213 EKGOAERKIKTLQ-TGSASTLGVEFLQASKLPWDQDVLMSFFFDALKEQVQOQLWEKD 271
QY 299 VAKSLSALIGOCIHIER---LARAARAKPRSPRALVLPHTIASHHVDPT--EPVGG 352
DB 272 RPTLVYINNAKIDDRQFAWTRNSRGNKRODNPRYHANOGRTQDTSYGTAGP 331
QY 353 ARMLTQEEKERRKLNLCYCGTGGHYADNC--PAKASK 390
DB 332 MATGMTKRDASK---VTCYNGCKKGHYERECKNPVKTNQ 367

RESULT 8
Q9HFY8 PRELIMINARY; PRT; 837 AA.
AC Q9HFY8;
DT 01-MAR-2001 (TREMBLrel. 16, Created)
DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE GAG PROTEIN.
GN GAG.
OS Colletotrichum gloeosporioides (Anthracnose fungus) (Glomerella
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
OC Sordariomycetes incertae sedis; Phyllachorales; Phyllachoraceae;
OC Glomerella.
OX NCBI_TaxID=5457;
RN [1]
RP SEQUENCE FROM N.A.
RC TRANSPOSON=RETROTRANSPOSON CGRET;
RA Zhu P., Oudemans P.V.;
RT "A long terminal repeat retrotransposon Cyret from the phytopathogenic
RT fungus Colletotrichum gloeosporioides on cranberry.";
RL Curr. Genet. 0:0-0(2000).
DR EMBL: AF264028; AAG24791.1; -.
DR InterPro: IPR001878; Znf_CCHC.
DR Pfam: PF00098; zf-CCHC; 1.
DR SMART: SM00343; Znf_C2HC; 1.
KW Zinc-finger.
SQ SEQUENCE 837 AA; 97738 MW; EF1D4BC70FD55003 CRC64;

Query Match 10.7%; Score 229.5; DB 3; Length 837;
Best Local Similarity 20.6%; Pred. No. 2.2e-09;
Matches 76; Conservative 68; Mismatches 174; Indels 51; Gaps 8;

QY 68 GHKSACVP-NWTERRRDELSEINNLR---EKVMKQSENNNLSQVOKLTEENTT----119
DB 13 GSSSRQVPQNLQDELQDKMKDDSDSEDETELKQLEKTNNDPKEMRDMFNQAAALKE 72

Appl. Environ. Microbiol. 61:1068-1072(1995).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=NRRL 1997; TRANSPOSON=AFRTL-1;
RA Okubara P.A., Tibbot B., McAlpin C.E., Hua S.-S.T.;
RT "AfRL-1, a retrotransposon-like element in the aflatoxin-producing
RT fungus Aspergillus flavus.";
RL Submitted (MAR-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF362957; AAL26311.1; -.
FT NON_TER 1252 1252
FT SEQUENCE 1252 AA; 144752 MW; D8FDC5DD8D8E6C4E8 CRC64;
SQ SEQUENCE 853 AA; 99422 MW; 1AEDD18CFBBA5B06 CRC64;

Query Match 13.0%; Score 279; DB 3; Length 1252;
Best Local Similarity 24.3%; Pred. No. 5.7e-13;
Matches 94; Conservative 72; Mismatches 155; Indels 66; Gaps 14;

QY 58 NNNNNKHTGHKS-----ACVPNMTERRRDELSEINNLEKVKMKQSENNNL 106
DB 2 SSQSSSKKTPVKSTPPAETDSESETTVKEQLKQMKSMITQLVNNAKEK---NQEIENL 57
QY 107 QSOVKLTEENTTLREQVEPTPEDEDDIELRGAAAAAAPPPI-BECCPELPEKFDGN 165
DB 58 KVQ-----LGEARTINEQDHIADQVQASAPKDAIGKVLKP--AEPDGT 105
QY 166 PDLAPFMAOQIPEKSTDFSDVRVRCVFTSMMTGRAARWASAKLERSHY--LMHNY 223
DB 106 RSKLQAFLTQNMHIHANRKNLIDEADKVFISTHLRGAANNPEPVI-REYIEVVPDNW 164
QY 224 PAFMEM-----KHV---FEDPQREVAKKIRRLROGMSGVIDYSNAFOMIAODLD 272
DB 165 SNTTREFLTDGSLNRKHLERTFGVDAAEVAERKLQLYQ-RGSASTYAAEFQIISRMD 223
QY 273 WNEPALIDQYHEGLSDHIQELSLEHVAKSLSALIGOCIHIERLARA-----AARK 325
DB 224 WNEGLCVNLHQWSQGHVQDEFAFRIDRATLNEAIDFAVKVDNRYHRLMEKRDNEAWRK 283
QY 326 PRSPRALVL-----PHIASHHQVDPTPEVGVGARMRLTQEEKERRKLNLC 371
DB 284 GSHRPKGQYKNDQRTGVKNDPYGPKPMELDTEGQGSK-GISQKERERRREKLC 342
QY 372 LYCGTGGHYADNCPAKASKSSPAGNSP 398
DB 343 YNCGRAGHMSKDCRQKRN-SOPANRKP 368

RESULT 7
Q00833 PRELIMINARY; PRT; 853 AA.
AC Q00833;
DT 01-NOV-1996 (TREMBLrel. 01, Created)
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE GAG POLYPROTEIN.
GN GAG.
OS Fusarium oxysporum.
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
OC Hypocreales; mitosporic Hypocreales; Fusarium.
OX NCBI_TaxID=5507;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=F. SP. LYCOPERSICI 42-87;
RX MEDLINE=96132549; PubMed=8544829;
RA Anaya N., Roncero M.I.;
RT "Skippy, a retrotransposon from the fungal plant pathogen Fusarium
RT oxysporum.";
RL Mol. Gen. Genet. 249:637-647(1995).
DR EMBL: L34658; AAA88790.1; -.
DR InterPro: IPR001878; Znf_CCHC.
DR Pfam: PF00098; zf-CCHC; 1.
DR PRINTS: PR00939; C2HC2NFINGER.
DR SMART: SM00343; Znf_C2HC; 1.
KW Polyprotein; Zinc-finger.
```



```

RN RP SEQUENCE FROM N.A.
RC STRAIN=NIPPONBARE;
RT "Rice genomic sequence.";
RL Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.
[2]
RN RP SEQUENCE FROM N.A.
RC STRAIN=NIPPONBARE;
RA Cordon H., Johnson D., Minx P., de la Bastide M., Nascimento L.U.,
RA Spiegel L.A., Kirchoff K.A., King L., Preston R.R., Vil M.D.,
RA Baker J.P., Miller B., Santos L., Zutavern T., Kuit K.H.,
RA Rodriguez S., Cunniss D.M., Balija V.S., Shah R.S., Bahret A., Bal H.,
RA O'Shaughnessy A., Dedhia N.N., McCombie W.R.;
RL Submitted (MAR-2001) to the EMBL/GenBank/DBJ databases.
[3]
RN RP SEQUENCE FROM N.A.
RC STRAIN=NIPPONBARE;
RA Palmer L.E., Bal H., Cordon H., Johnson D., Minx P., de la Bastide M.,
RA Nascimento L.U., Spiegel L.A., Kirchoff K.A., King L., Preston R.R.,
RA Vil M.D., Baker J.P., Miller B., Santos L., Zutavern T., Kuit K.H.,
RA Rodriguez S., Cunniss D.M., Balija V.S., Shah R.S., Bahret A.,
RA O'Shaughnessy A., Dedhia N.N., McCombie W.R.;
RL Submitted (MAY-2001) to the EMBL/GenBank/DBJ databases.
[4]
RN RP SEQUENCE FROM N.A.
RC STRAIN=NIPPONBARE;
RA Palmer L.E., Bal H., Cordon H., Johnson D., Minx P., de la Bastide M.,
RA Nascimento L.U., Spiegel L.A., Kirchoff K.A., King L., Preston R.R.,
RA Vil M.D., Baker J.P., Miller B., Santos L., Zutavern T., Kuit K.H.,
RA Rodriguez S., Cunniss D.M., Balija V.S., Shah R.S., Bahret A.,
RA O'Shaughnessy A., Dedhia N.N., McCombie W.R.;
RL Submitted (AUG-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL: AC022352; AAK51574.1; -.
SQ SEQUENCE 1810 AA; 205746 MW; 5AB92CD640BAB1F2 CRC64;

Query Match 8.8%; Score 189; DB 10; Length 1810;
Best Local Similarity 20.7%; Pred. No. 7.6e-06;
Matches 100; Conservative 57; Mismatches 125; Indels 202; Gaps 24;

QY 5 RVLTKKR-RSGRGQDPGLPHRSEATAGRSPTPTVTILGDCPPPPPPPPNNNNNN 63
DB 304 QMVTTRNVNTGEGNPEGSNNHQ-----GNPPPP-----PPPPPPPPDTN---- 346
QY 64 SKHTGHSACVNPNTERRRDELSEENLNREKVMKQSENNLNQSQVKLTTEENTTLREQ 123
DB 347 -----ALTOI-----LAQANMM-----TAFLHH 366
QY 124 VEPTPEDDDDIELRGAAGAAAAPPPPIEECPEDL---PEKF--DGNP-DMLAPFMA--- 174
DB 367 LQNPPOON-----APPPPOHSLAEFLRIRPTTFSSNNPVDALDHLHVGK 414
QY 175 -----QCQIFMEKSTRDFSVDVRVCFVTSMTGRAARW-----ASA 211
DB 415 KLDTVQCS-----DEEKVFAAHOLQGPASLMDHFOATQPEGQPTITWRTFA 462
QY 212 KLSHYLMHNYPAFMMEMKHVFEDPQRRVAKRKIRRLROGMSGVSDYSNAFOMIA--- 268
DB 463 AFRTH-----VPAGVVALK-----KREFELKQGNRSWMEYLHEFNNLARYA 505
QY 269 -QLDWNNEPALIDQYHEGLSDHIQELSHLEVA---KSLSALIGQCIIHE----- 314
DB 506 PEDVREDE-----EKQEKFLAGMDPELSVRLVSGDYPDFQRLVDKSLRLEAKHKELES 560
QY 315 RLIA-----RAAARPRSPRALVPLPHIASHHOVD----- 345
DB 561 RLANFRNQOGANQVRVYTNYPYGGSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSS 620
QY 346 -----PT-EPVGGARMRLTOEEKERRRKLNLCLYCGTGGHYADNCP---AKAKSSP 393
DB 621 TRAPRPTTVPVQGR-----DAQOQRLCFCNCFEPGHFADKCKPRRQOQAPP 672
```

```

QY 394 AGNS 397
DB 673 RSNN 676

RESULT 12
Q94DA0 PRELIMINARY; PRT; 1500 AA.
AC Q94DA0;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE PUTATIVE POLYPROTEIN.
GN P0439E11.5.
OS Oryza sativa (Rice).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC Ehrhartoideae; Oryzaceae; Oryza.
OX NCBI_TaxID=4530;
RN [1]
RC STRAIN=CV. NIPPONBARE;
RA Sasaki T., Matsumoto T., Yamamoto K.;
RT "Oryza sativa nipponbare(GA3) genomic DNA, chromosome 1, PAC
clone:P0439E11.";
RL Submitted (FEB-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL: AP003315; BAB61255.1; -.
KW Polyprotein.
SQ SEQUENCE 1500 AA; 171517 MW; 5DE27E21F9ABE861 CRC64;

Query Match 8.6%; Score 185; DB 10; Length 1500;
Best Local Similarity 20.3%; Pred. No. 1.2e-05;
Matches 98; Conservative 53; Mismatches 125; Indels 206; Gaps 21;

QY 6 VLTKKR-RSGRGQDPGLPHRSEATAGRSPTPTVTILGDCPPPPPPPPNNNNNNNS 64
DB 1 MVSTRNVNTGEGNPEGSNPN-----POGNPPPPPPPPDTN----- 37
QY 65 KHTGHSACVNPNTERRRDELSEENLNREKVMKQSENNLNQSQVKLTTEENTTLREQ 124
DB 38 -----ALTOIQAQANMMNAFLHHLQNLPOON----- 65
QY 125 EPTPEDDDDIELRGAAGAAAAPPPPIEECPEDL---PEKF--DGNP-DMLAPFMA--- 174
DB 66 -----APPPPOHSLAEFLRIRPTTFSSNNPVDALDHLHVGK 106
QY 175 -----QCQIFMEKSTRDFSVDVRVCFVTSMTGRAARW-----ASAK 212
DB 107 LDTVQCS-----DEEKVFAAHOLQGPASLMDHFOATQPEGQPTITWARTAA 154
QY 213 LERSHYLMHNYPAFMMEMKHVFEDPQRRVAKRKIRRLROGMSGVSDYSNAFOMIA--- 268
DB 155 FRTH-----VPAGVVALK-----KREFELKQGNRSWMEYLHEFNNLARYAP 197
QY 269 QLDWNNEPALIDQYHEGLSDHIQELSHLEVA---KSLSALIGQCIIHE-----R 315
DB 198 EDVREDE-----EKQEKFLAGMDPELSVRLVSGDYPDFQRLVDKSLRLEAKHKELES 252
QY 316 RLIA-----RAAARPRSPRALVPLPHIASHHOVD----- 345
DB 253 RLANFRNQOGANQVRVYTNYPYGGSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSS 312
QY 346 -----PT-EPVGGARMRLTOEEKERRRKLNLCLYCGTGGHYADNCP---AKAKSSPAG 395
DB 313 APRPPTTVPVQGR-----DAQOQRLCFCNCFEPGHFADKCKPRRQOQAPP 364
QY 396 NS 397
DB 365 NN 366

RESULT 13
```







---

**THIS PAGE BLANK (USPTO)**

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 13, 2002, 15:11:14 ; Search time 31.31 Seconds  
(without alignments)  
1422.568 Million cell updates/sec

Title: US-09-631-863A-2  
Perfect score: 401  
Sequence: 1 MNKRVLTKKRRSGRGQD.....DNCPAKAKSSPAGNSAPL 401

Scoring table: OLIGO  
Gapop 60.0 , Gapext 60.0

Searched: 747574 seqs, 111073796 residues

Word size : 0

Total number of hits satisfying chosen parameters: 747574

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : A\_Geneseq\_032802.\*  
1: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1980.DAT.\*  
2: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1981.DAT.\*  
3: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1982.DAT.\*  
4: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1983.DAT.\*  
5: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1984.DAT.\*  
6: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1985.DAT.\*  
7: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1986.DAT.\*  
8: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1987.DAT.\*  
9: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1988.DAT.\*  
10: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1989.DAT.\*  
11: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1990.DAT.\*  
12: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1991.DAT.\*  
13: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1992.DAT.\*  
14: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1993.DAT.\*  
15: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1994.DAT.\*  
16: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1995.DAT.\*  
17: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1996.DAT.\*  
18: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1997.DAT.\*  
19: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1998.DAT.\*  
20: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1999.DAT.\*  
21: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA2000.DAT.\*  
22: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA2001.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	ID	Description
1	342	85.3	342	21	Human ORFX ORF1912
2	10	2.5	174	21	Arabidopsis thalia
3	10	2.5	209	21	Arabidopsis thalia
4	10	2.5	235	21	Arabidopsis thalia
5	10	2.5	235	21	Arabidopsis thalia
6	10	2.5	237	21	Arabidopsis thalia
7	10	2.5	237	21	Arabidopsis thalia
8	10	2.5	237	21	Arabidopsis thalia
9	10	2.5	240	21	Arabidopsis thalia
10	10	2.5	240	21	Arabidopsis thalia
11	10	2.5	240	21	Arabidopsis thalia

12	10	2.5	263	21	AAG39267	Arabidopsis thalia
13	10	2.5	697	22	ABG25300	Novel human diagno
14	9	2.2	9	3	AAP20304	Peptide for use in
15	9	2.2	10	22	AAU09969	Proline linker seq
16	9	2.2	11	20	AAW98980	FAM-(Pro)10-Lys(ep
17	9	2.2	13	20	AAW98981	FAM-(Pro)12-Lys(ep
18	9	2.2	13	22	AAB30651	Trypsin modulating
19	9	2.2	13	22	AAB30652	Trypsin modulating
20	9	2.2	15	18	AAW38951	Peptide resembling
21	9	2.2	15	18	AAW58137	Gonadotropin relea
22	9	2.2	15	22	AAB30653	Trypsin modulating
23	9	2.2	17	18	AAW39014	Peptide resembling
24	9	2.2	18	18	AAW39010	Peptide resembling
25	9	2.2	18	18	AAW38974	Peptide resembling
26	9	2.2	18	22	AAB30654	Trypsin modulating
27	9	2.2	19	18	AAW39009	Peptide resembling
28	9	2.2	19	22	AAB30655	Trypsin modulating
29	9	2.2	20	18	AAW38960	Peptide resembling
30	9	2.2	20	18	AAW38961	Peptide resembling
31	9	2.2	20	18	AAW38962	Peptide resembling
32	9	2.2	20	18	AAW38971	Peptide resembling
33	9	2.2	21	22	AAB30656	Trypsin modulating
34	9	2.2	23	18	AAW38979	Peptide resembling
35	9	2.2	23	22	AAB30657	Trypsin modulating
36	9	2.2	26	22	AAO04983	Human polypeptide
37	9	2.2	29	22	AAO04677	Human polypeptide
38	9	2.2	33	21	AAV54915	RGD-containing syn
39	9	2.2	33	22	AAO04814	Human polypeptide
40	9	2.2	34	5	AAP40632	Sequence of sperma
41	9	2.2	39	22	ABG27937	Novel human diagno
42	9	2.2	41	22	AAO02061	Human polypeptide
43	9	2.2	42	22	ABG16321	Novel human diagno
44	9	2.2	44	22	AAO04929	Human polypeptide
45	9	2.2	46	22	ABG12808	Novel human diagno

ALIGNMENTS

RESULT 1  
AAB42148  
ID AAB42148 standard; Protein; 342 AA.  
AC AAB42148;  
AC AAB42148;  
DT 08-FEB-2001 (first entry)  
XX Human ORFX ORF1912 polypeptide sequence SEQ ID NO:3824.

Human; open reading frame; ORFX; detection; cytotstatic; hepatotropic;  
vulnery; antipsoriatic; antiparkinsonian; nootropic; neuroprotective;  
anticonvulsant; osteopathic; antiarthritic; immunosuppressant; cardiac;  
immunostimulant; thrombolytic; coagulant; vasotropic; antidiabetic;  
KW hypotensive; dermatological; immunosuppressive; antiinflammatory;  
KW antiviral; antibacterial; antifungal; antirheumatic; antithyroid;  
KW antianemic; gene therapy; cancer; proliferative disorder; hypertension;  
KW neurodegenerative disorder; osteoarthritis; graft vs host disease;  
KW cardiovascular disease; diabetes mellitus; hypothyroidism; SCID; AIDS;  
KW cholesterol ester storage; systemic lupus erythematosus; infection;  
KW severe combined immunodeficiency; malaria; autoimmune disorder; asthma;  
KW allergy; aplastic anaemia; nocturnal haemoglobinuria; burn; wound;  
KW bone damage; cartilage damage; antiinflammatory disease; coagulation;  
KW thrombosis; contraceptive.

XX Homo sapiens.

XX WO200058473-A2.

XX 05-OCT-2000.

XX 31-MAR-2000; 2000WO-US08621.

XX 31-MAR-1999; 99US-0127607.

PR 02-APR-1999; 99US-0127636.  
PR 05-APR-1999; 99US-0127728.  
PR 30-MAR-2000; 2000US-0540763.  
XX  
PA (CURA-) CURAGEN CORP.  
XX  
XX Shimkets RA, Leach M;  
XX  
XX WPI; 2000-602362/57.  
DR N-PSDB; AAC76357.  
XX  
XX Novel nucleic acids and peptides derived from open reading frame X,  
PT useful for treating e.g. cancers, proliferative disorders,  
PT neurodegenerative disorders and cardiovascular disease -  
XX  
XX Claim 11; Page 2977-2978; 5507pp; English.  
XX  
XX AAC74446 to AAC77606 encode the proteins given in AAB40237 to AAB43397,  
CC which represent the human ORFX open reading frames 1 to 3161. The ORFX  
CC sequences have activities such as: cytostatic; hepatotropic; vulnary;  
CC antiposiatric; antiparkinsonian; nootropic; neuroprotective;  
CC osteopathic; anticonvulsant; antiarthritic; immunosuppressant;  
CC immunostimulant; cardiac; thrombolytic; coagulant; vasotropic;  
CC antidiabetic; hypotensive; dermatological; immunosuppressive;  
CC antiinflammatory; antibacterial; antiviral; antifungal; antirheumatic;  
CC antithyroid; and antianemic. The sequences can be used for determining  
CC the presence of or predisposition to, or preventing or treating  
CC pathological conditions associated with an ORFX-associated disorder. The  
CC nucleic acids can be used to express ORFX proteins in gene therapy  
CC vectors. The proteins and nucleic acids may be used to treat cancers,  
CC proliferative disorders, neurodegenerative disorders, osteoarthritis,  
CC graft vs host disease, cardiovascular disease, diabetes mellitus,  
CC hypertension, hypothyroidism, cholesterol ester storage, systemic lupus  
CC erythematosus, severe combined immunodeficiency (SCID), AIDS, viral,  
CC bacterial or fungal infection, malaria, autoimmune disorders, asthma,  
CC allergies, aplastic anemia, burns, wounds, bone and cartilage damage,  
CC nocturnal haemoglobinuria, antiinflammatory disease; to enhance  
CC coagulation; to inhibit thrombosis; and as a contraceptive.  
XX  
SQ Sequence 342 AA;

Query Match 85.3%; Score 342; DB 21; Length 342;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 342; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 60 NNNKKTGKHSACVPNMTERRRDELSEENNLREKVMKQSENNNLQSQVQKLTTENTT 119  
Db 1 nnnskktghksacvpnmterrrdelseeinnlrekvmkqseennnlqsqvqkltteentt 60  
QY 120 LREQVEPTPEDEDDIELRGAAAAAAPPPTIEECPEDLPEKFGNPDMLAPFMAQCQIF 179  
Db 61 lreqvptpedddielrgaaaaaappppieecpedlpekfgnpgdmlapfmaqcf 120  
QY 180 MEKTRDFSDVRVRCVFTSMMTGRAARWASAKLERSHYLMHNYPAFMMEMKHVFEDPQR 239  
Db 121 mektrdfsvdrvrcvftsmmtgraarwasaklershylmhnypatfmmemkhvfedpqr 180  
QY 240 REVAKRIRLRQGMGSVIDYSNAFQMAQDLWNEPALIDQYHEGLSDHIQELSHLEV 299  
Db 181 revakrirlrqgmgsvidysnafqmaqlqdlwnepalidqyheglshleishlev 240  
QY 300 AKLSALIGQCIHTERRLARAARAKRPSPPRALVLPFIASHHQVDPTPEVGGARMRLTQ 359  
Db 241 akslsaligqcihierlaraaarkprspralvlpfiashhqvdpptevvgarmrltq 300  
QY 360 EEKERRKRLNLCLYCGGGHYADNCAPAKSSPAGNSPAPL 401  
Db 301 eekerrrklnlclycgggghyadncapaksspagnspapl 342

ID AAG18590 standard; Protein; 174 AA.  
XX AC AAG18590;  
XX DT 17-OCT-2000 (first entry)  
XX DE Arabidopsis thaliana protein fragment SEQ ID NO: 20061.  
XX KW Protein identification; signal transduction pathway; metabolic pathway;  
KW hybridisation assay; genetic mapping; gene expression control; promoter;  
KW termination sequence.  
XX OS Arabidopsis thaliana.  
XX PN EP1033405-A2.  
XX PD 06-SEP-2000.  
XX 25-FEB-2000; 2000EP-0301439.  
XX 25-FEB-1999; 99US-0121825.  
PR 05-MAR-1999; 99US-0123180.  
PR 09-MAR-1999; 99US-0123548.  
PR 23-MAR-1999; 99US-0125788.  
PR 25-MAR-1999; 99US-0126264.  
PR 29-MAR-1999; 99US-0126785.  
PR 01-APR-1999; 99US-0127462.  
PR 06-APR-1999; 99US-0128234.  
PR 08-APR-1999; 99US-0128714.  
PR 16-APR-1999; 99US-0129845.  
PR 19-APR-1999; 99US-0130077.  
PR 21-APR-1999; 99US-0130449.  
PR 23-APR-1999; 99US-0130510.  
PR 23-APR-1999; 99US-0130891.  
PR 28-APR-1999; 99US-0131449.  
PR 30-APR-1999; 99US-0132048.  
PR 30-APR-1999; 99US-0132407.  
PR 04-MAY-1999; 99US-0132484.  
PR 05-MAY-1999; 99US-0132485.  
PR 06-MAY-1999; 99US-0132486.  
PR 06-MAY-1999; 99US-0132487.  
PR 07-MAY-1999; 99US-0132863.  
PR 11-MAY-1999; 99US-0134256.  
PR 14-MAY-1999; 99US-0134218.  
PR 14-MAY-1999; 99US-0134219.  
PR 14-MAY-1999; 99US-0134221.  
PR 18-MAY-1999; 99US-0134370.  
PR 18-MAY-1999; 99US-0134768.  
PR 19-MAY-1999; 99US-0134941.  
PR 20-MAY-1999; 99US-0135124.  
PR 21-MAY-1999; 99US-0135353.  
PR 24-MAY-1999; 99US-0135629.  
PR 25-MAY-1999; 99US-0136021.  
PR 27-MAY-1999; 99US-0136392.  
PR 28-MAY-1999; 99US-0136782.  
PR 01-JUN-1999; 99US-0137222.  
PR 03-JUN-1999; 99US-0137528.  
PR 04-JUN-1999; 99US-0137502.  
PR 07-JUN-1999; 99US-0137724.  
PR 08-JUN-1999; 99US-0138094.  
PR 10-JUN-1999; 99US-0138540.  
PR 10-JUN-1999; 99US-0138847.  
PR 14-JUN-1999; 99US-0139119.  
PR 16-JUN-1999; 99US-0139452.  
PR 16-JUN-1999; 99US-0139453.  
PR 17-JUN-1999; 99US-0139492.  
PR 18-JUN-1999; 99US-0139454.  
PR 18-JUN-1999; 99US-0139455.  
PR 18-JUN-1999; 99US-0139456.  
PR 18-JUN-1999; 99US-0139457.  
PR 18-JUN-1999; 99US-0139458.  
PR 18-JUN-1999; 99US-0139459.  
PR 18-JUN-1999; 99US-0139460.

PR 18-JUN-1999; 99US-0139461.  
PR 18-JUN-1999; 99US-0139462.  
PR 18-JUN-1999; 99US-0139463.  
PR 18-JUN-1999; 99US-0139750.  
PR 18-JUN-1999; 99US-0139763.  
PR 21-JUN-1999; 99US-0139817.  
PR 22-JUN-1999; 99US-0139899.  
PR 23-JUN-1999; 99US-0140353.  
PR 23-JUN-1999; 99US-0140354.  
PR 24-JUN-1999; 99US-0140695.  
PR 28-JUN-1999; 99US-0140823.  
PR 29-JUN-1999; 99US-0140991.  
PR 30-JUN-1999; 99US-0141287.  
PR 01-JUL-1999; 99US-0141842.  
PR 01-JUL-1999; 99US-0142154.  
PR 02-JUL-1999; 99US-0142055.  
PR 06-JUL-1999; 99US-0142390.  
PR 08-JUL-1999; 99US-0142803.  
PR 09-JUL-1999; 99US-0142920.  
PR 12-JUL-1999; 99US-0142977.  
PR 13-JUL-1999; 99US-0143542.  
PR 14-JUL-1999; 99US-0143624.  
PR 15-JUL-1999; 99US-0144005.  
PR 16-JUL-1999; 99US-0144085.  
PR 16-JUL-1999; 99US-0144086.  
PR 19-JUL-1999; 99US-0144325.  
PR 19-JUL-1999; 99US-0144331.  
PR 19-JUL-1999; 99US-0144332.  
PR 19-JUL-1999; 99US-0144333.  
PR 19-JUL-1999; 99US-0144334.  
PR 19-JUL-1999; 99US-0144335.  
PR 20-JUL-1999; 99US-0144352.  
PR 20-JUL-1999; 99US-0144632.  
PR 20-JUL-1999; 99US-0144632.  
PR 21-JUL-1999; 99US-0144884.  
PR 21-JUL-1999; 99US-0144814.  
PR 21-JUL-1999; 99US-0145086.  
PR 21-JUL-1999; 99US-0145088.  
PR 22-JUL-1999; 99US-0145085.  
PR 22-JUL-1999; 99US-0145087.  
PR 22-JUL-1999; 99US-0145089.  
PR 22-JUL-1999; 99US-0145192.  
PR 23-JUL-1999; 99US-0145145.  
PR 23-JUL-1999; 99US-0145218.  
PR 23-JUL-1999; 99US-0145224.  
PR 26-JUL-1999; 99US-0145276.  
PR 27-JUL-1999; 99US-0145913.  
PR 27-JUL-1999; 99US-0145918.  
PR 27-JUL-1999; 99US-0145919.  
PR 28-JUL-1999; 99US-0145951.  
PR 02-AUG-1999; 99US-0146386.  
PR 02-AUG-1999; 99US-0146388.  
PR 02-AUG-1999; 99US-0146389.  
PR 03-AUG-1999; 99US-0147038.  
PR 04-AUG-1999; 99US-0147204.  
PR 04-AUG-1999; 99US-0147302.  
PR 05-AUG-1999; 99US-0147192.  
PR 05-AUG-1999; 99US-0147260.  
PR 06-AUG-1999; 99US-0147303.  
PR 06-AUG-1999; 99US-0147416.  
PR 09-AUG-1999; 99US-0147935.  
PR 09-AUG-1999; 99US-0147935.  
PR 10-AUG-1999; 99US-0148171.  
PR 11-AUG-1999; 99US-0148319.  
PR 12-AUG-1999; 99US-0148341.  
PR 13-AUG-1999; 99US-0148565.  
PR 13-AUG-1999; 99US-0148684.  
PR 16-AUG-1999; 99US-0149368.  
PR 17-AUG-1999; 99US-0149175.  
PR 18-AUG-1999; 99US-0149426.  
PR 20-AUG-1999; 99US-0149722.  
PR 20-AUG-1999; 99US-0149723.  
PR 20-AUG-1999; 99US-0149929.  
PR 23-AUG-1999; 99US-0149902.

PR 23-AUG-1999; 99US-0149930.  
PR 25-AUG-1999; 99US-0150566.  
PR 26-AUG-1999; 99US-0150884.  
PR 27-AUG-1999; 99US-0151065.  
PR 27-AUG-1999; 99US-0151066.  
PR 27-AUG-1999; 99US-0151080.  
PR 30-AUG-1999; 99US-0151303.  
PR 31-AUG-1999; 99US-0151438.  
PR 01-SEP-1999; 99US-0151930.  
PR 07-SEP-1999; 99US-0152363.  
PR 10-SEP-1999; 99US-0153070.  
PR 13-SEP-1999; 99US-0153758.  
PR 15-SEP-1999; 99US-0154018.  
PR 16-SEP-1999; 99US-0154039.  
PR 20-SEP-1999; 99US-0154779.  
PR 22-SEP-1999; 99US-0155139.  
PR 23-SEP-1999; 99US-0155486.  
PR 24-SEP-1999; 99US-0155659.  
PR 28-SEP-1999; 99US-0156458.  
PR 29-SEP-1999; 99US-0156596.  
PR 04-OCT-1999; 99US-0157117.  
PR 05-OCT-1999; 99US-0157753.  
PR 06-OCT-1999; 99US-0157865.  
PR 07-OCT-1999; 99US-0158029.  
PR 08-OCT-1999; 99US-0158232.  
PR 12-OCT-1999; 99US-0158369.  
PR 13-OCT-1999; 99US-0159293.  
PR 13-OCT-1999; 99US-0159294.  
PR 13-OCT-1999; 99US-0159295.  
PR 14-OCT-1999; 99US-0159329.  
PR 14-OCT-1999; 99US-0159330.  
PR 14-OCT-1999; 99US-0159331.  
PR 14-OCT-1999; 99US-0159637.  
PR 14-OCT-1999; 99US-0159638.  
PR 18-OCT-1999; 99US-0159584.  
PR 21-OCT-1999; 99US-0160741.  
PR 21-OCT-1999; 99US-0160767.  
PR 21-OCT-1999; 99US-0160768.  
PR 21-OCT-1999; 99US-0160770.  
PR 21-OCT-1999; 99US-0160814.  
PR 21-OCT-1999; 99US-0160815.  
PR 22-OCT-1999; 99US-0160980.  
PR 22-OCT-1999; 99US-0160981.  
PR 22-OCT-1999; 99US-0160989.  
PR 25-OCT-1999; 99US-0161404.  
PR 25-OCT-1999; 99US-0161405.  
PR 25-OCT-1999; 99US-0161406.  
PR 26-OCT-1999; 99US-0161359.  
PR 26-OCT-1999; 99US-0161360.  
PR 26-OCT-1999; 99US-0161361.  
PR 28-OCT-1999; 99US-0161920.  
PR 28-OCT-1999; 99US-0161992.  
PR 28-OCT-1999; 99US-0161993.  
PR 29-OCT-1999; 99US-0162142.

Query Match 2.5%; Score 10; DB 21; Length 174;  
Best Local Similarity 100.0%; Pred. No. 0.4;

Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 48 pppppppppn 57  
| | | | | | | | | |  
Db 18 pppppppppn 27

RESULT 3  
AAG18589  
ID AAG18589 standard; Protein; 209 AA.  
XX  
AC AAG18589;  
XX  
DT 17-OCT-2000 (first entry)  
XX  
DE Arabidopsis thaliana protein fragment SEQ ID NO: 20060.

XX Protein identification; signal transduction pathway; metabolic pathway;  
KW hybridisation assay; genetic mapping; gene expression control; promoter;  
KW termination sequence.  
XX  
OS Arabidopsis thaliana.  
XX  
PN EP1033405-A2.  
XX  
PD 06-SEP-2000.  
XX  
PF 25-FEB-2000; 2000EP-0301439.  
XX  
PR 25-FEB-1999; 99US-0121825.  
PR 05-MAR-1999; 99US-0123180.  
PR 09-MAR-1999; 99US-0123548.  
PR 23-MAR-1999; 99US-0125788.  
PR 25-MAR-1999; 99US-0126264.  
PR 29-MAR-1999; 99US-0126785.  
PR 01-APR-1999; 99US-0127462.  
PR 06-APR-1999; 99US-0128234.  
PR 08-APR-1999; 99US-0128714.  
PR 16-APR-1999; 99US-0129845.  
PR 19-APR-1999; 99US-0130077.  
PR 21-APR-1999; 99US-0130449.  
PR 23-APR-1999; 99US-0130510.  
PR 28-APR-1999; 99US-0130891.  
PR 30-APR-1999; 99US-0131449.  
PR 30-APR-1999; 99US-0132048.  
PR 30-APR-1999; 99US-0132407.  
PR 04-MAY-1999; 99US-0132484.  
PR 05-MAY-1999; 99US-0132485.  
PR 06-MAY-1999; 99US-0132486.  
PR 06-MAY-1999; 99US-0132487.  
PR 07-MAY-1999; 99US-0132863.  
PR 11-MAY-1999; 99US-0134256.  
PR 14-MAY-1999; 99US-0134218.  
PR 14-MAY-1999; 99US-0134219.  
PR 14-MAY-1999; 99US-0134221.  
PR 14-MAY-1999; 99US-0134370.  
PR 18-MAY-1999; 99US-0134768.  
PR 19-MAY-1999; 99US-0134941.  
PR 20-MAY-1999; 99US-0135124.  
PR 21-MAY-1999; 99US-0135353.  
PR 24-MAY-1999; 99US-0135629.  
PR 25-MAY-1999; 99US-0136021.  
PR 27-MAY-1999; 99US-0136392.  
PR 28-MAY-1999; 99US-0136782.  
PR 01-JUN-1999; 99US-0137222.  
PR 03-JUN-1999; 99US-0137528.  
PR 04-JUN-1999; 99US-0137502.  
PR 07-JUN-1999; 99US-0137724.  
PR 08-JUN-1999; 99US-0138094.  
PR 10-JUN-1999; 99US-0138540.  
PR 10-JUN-1999; 99US-0138847.  
PR 14-JUN-1999; 99US-0139119.  
PR 16-JUN-1999; 99US-0139452.  
PR 16-JUN-1999; 99US-0139453.  
PR 17-JUN-1999; 99US-0139492.  
PR 18-JUN-1999; 99US-0139454.  
PR 18-JUN-1999; 99US-0139455.  
PR 18-JUN-1999; 99US-0139456.  
PR 18-JUN-1999; 99US-0139457.  
PR 18-JUN-1999; 99US-0139458.  
PR 18-JUN-1999; 99US-0139459.  
PR 18-JUN-1999; 99US-0139460.  
PR 18-JUN-1999; 99US-0139461.  
PR 18-JUN-1999; 99US-0139462.  
PR 18-JUN-1999; 99US-0139463.  
PR 18-JUN-1999; 99US-0139750.  
PR 18-JUN-1999; 99US-0139763.  
PR 21-JUN-1999; 99US-0139817.  
PR 22-JUN-1999; 99US-0139899.  
PR 23-JUN-1999; 99US-0140353.  
PR 23-JUN-1999; 99US-0140354.  
PR 24-JUN-1999; 99US-0140695.  
PR 28-JUN-1999; 99US-0140823.  
PR 29-JUN-1999; 99US-0140991.  
PR 30-JUN-1999; 99US-0141287.  
PR 01-JUL-1999; 99US-0141842.  
PR 01-JUL-1999; 99US-0142154.  
PR 02-JUL-1999; 99US-0142055.  
PR 06-JUL-1999; 99US-0142390.  
PR 08-JUL-1999; 99US-0142803.  
PR 09-JUL-1999; 99US-0142920.  
PR 12-JUL-1999; 99US-0142977.  
PR 13-JUL-1999; 99US-0143542.  
PR 14-JUL-1999; 99US-0143624.  
PR 15-JUL-1999; 99US-0144005.  
PR 16-JUL-1999; 99US-0144085.  
PR 16-JUL-1999; 99US-0144086.  
PR 19-JUL-1999; 99US-0144325.  
PR 19-JUL-1999; 99US-0144331.  
PR 19-JUL-1999; 99US-0144332.  
PR 19-JUL-1999; 99US-0144333.  
PR 19-JUL-1999; 99US-0144334.  
PR 19-JUL-1999; 99US-0144335.  
PR 20-JUL-1999; 99US-0144352.  
PR 20-JUL-1999; 99US-0144632.  
PR 20-JUL-1999; 99US-0144884.  
PR 21-JUL-1999; 99US-0144814.  
PR 21-JUL-1999; 99US-0145086.  
PR 21-JUL-1999; 99US-0145088.  
PR 22-JUL-1999; 99US-0145085.  
PR 22-JUL-1999; 99US-0145087.  
PR 22-JUL-1999; 99US-0145089.  
PR 22-JUL-1999; 99US-0145192.  
PR 23-JUL-1999; 99US-0145145.  
PR 23-JUL-1999; 99US-0145218.  
PR 23-JUL-1999; 99US-0145224.  
PR 26-JUL-1999; 99US-0145276.  
PR 27-JUL-1999; 99US-0145913.  
PR 27-JUL-1999; 99US-0145918.  
PR 27-JUL-1999; 99US-0145919.  
PR 28-JUL-1999; 99US-0145951.  
PR 02-AUG-1999; 99US-0146386.  
PR 02-AUG-1999; 99US-0146388.  
PR 02-AUG-1999; 99US-0146389.  
PR 03-AUG-1999; 99US-0147038.  
PR 04-AUG-1999; 99US-0147204.  
PR 04-AUG-1999; 99US-0147302.  
PR 05-AUG-1999; 99US-0147192.  
PR 05-AUG-1999; 99US-0147260.  
PR 06-AUG-1999; 99US-0147303.  
PR 06-AUG-1999; 99US-0147416.  
PR 09-AUG-1999; 99US-0147493.  
PR 09-AUG-1999; 99US-0147935.  
PR 10-AUG-1999; 99US-0148171.  
PR 11-AUG-1999; 99US-0148319.  
PR 12-AUG-1999; 99US-0148341.  
PR 13-AUG-1999; 99US-0148565.  
PR 13-AUG-1999; 99US-0148684.  
PR 16-AUG-1999; 99US-0149368.  
PR 17-AUG-1999; 99US-0149175.  
PR 18-AUG-1999; 99US-0149426.  
PR 20-AUG-1999; 99US-0149722.  
PR 20-AUG-1999; 99US-0149723.  
PR 20-AUG-1999; 99US-0149929.  
PR 23-AUG-1999; 99US-0149902.  
PR 23-AUG-1999; 99US-0149930.  
PR 25-AUG-1999; 99US-0150566.  
PR 26-AUG-1999; 99US-0150884.  
PR 27-AUG-1999; 99US-0151065.  
PR 27-AUG-1999; 99US-0151066.  
PR 27-AUG-1999; 99US-0151080.  
PR 30-AUG-1999; 99US-0151303.

```
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.
```

Query Match Best Local Similarity 2.5%; Score 10; DB 21; Length 209;

Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy 48 pppppppppn 57
    |iiiiiii
Db 53 pppppppppn 62
```

RESULT 4

AAG11272 AAG11272 standard; Protein; 235 AA.

XX AC AAG11272;

XX DT 17-OCT-2000 (first entry)

XX DE Arabidopsis thaliana protein fragment SEQ ID NO: 9926.

XX KW Protein identification; signal transduction pathway; metabolic pathway;

XX KW hybridisation assay; genetic mapping; gene expression control; promoter;

XX KW termination sequence.

XX OS Arabidopsis thaliana.

```
PN EP1033405-A2.
XX 06-SEP-2000.
XX 25-FEB-2000; 2000EP-0301439.
XX 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
```

PR 01-JUL-1999; 9905-0142154.  
PR 02-JUL-1999; 9905-0142055.  
PR 06-JUL-1999; 9905-0142390.  
PR 08-JUL-1999; 9905-0142803.  
PR 09-JUL-1999; 9905-0142920.  
PR 12-JUL-1999; 9905-0142977.  
PR 13-JUL-1999; 9905-0143542.  
PR 14-JUL-1999; 9905-0143624.  
PR 15-JUL-1999; 9905-0144005.  
PR 16-JUL-1999; 9905-0144085.  
PR 18-JUL-1999; 9905-0144086.  
PR 19-JUL-1999; 9905-0144325.  
PR 19-JUL-1999; 9905-0144331.  
PR 19-JUL-1999; 9905-0144332.  
PR 19-JUL-1999; 9905-0144333.  
PR 19-JUL-1999; 9905-0144334.  
PR 19-JUL-1999; 9905-0144335.  
PR 20-JUL-1999; 9905-0144352.  
PR 20-JUL-1999; 9905-0144632.  
PR 20-JUL-1999; 9905-0144884.  
PR 21-JUL-1999; 9905-0144814.  
PR 21-JUL-1999; 9905-0145086.  
PR 21-JUL-1999; 9905-0145088.  
PR 22-JUL-1999; 9905-0145085.  
PR 22-JUL-1999; 9905-0145087.  
PR 22-JUL-1999; 9905-0145089.  
PR 22-JUL-1999; 9905-0145192.  
PR 23-JUL-1999; 9905-0145145.  
PR 23-JUL-1999; 9905-0145218.  
PR 23-JUL-1999; 9905-0145224.  
PR 26-JUL-1999; 9905-0145276.  
PR 27-JUL-1999; 9905-0145913.  
PR 27-JUL-1999; 9905-0145918.  
PR 27-JUL-1999; 9905-0145919.  
PR 28-JUL-1999; 9905-0145951.  
PR 02-AUG-1999; 9905-0146386.  
PR 02-AUG-1999; 9905-0146388.  
PR 03-AUG-1999; 9905-0146389.  
PR 03-AUG-1999; 9905-0147038.  
PR 04-AUG-1999; 9905-0147204.  
PR 04-AUG-1999; 9905-0147302.  
PR 05-AUG-1999; 9905-0147192.  
PR 05-AUG-1999; 9905-0147260.  
PR 06-AUG-1999; 9905-0147303.  
PR 06-AUG-1999; 9905-0147416.  
PR 09-AUG-1999; 9905-0147493.  
PR 09-AUG-1999; 9905-0147935.  
PR 10-AUG-1999; 9905-0148171.  
PR 11-AUG-1999; 9905-0148319.  
PR 12-AUG-1999; 9905-0148341.  
PR 13-AUG-1999; 9905-0148565.  
PR 13-AUG-1999; 9905-0148684.  
PR 16-AUG-1999; 9905-0149368.  
PR 17-AUG-1999; 9905-0149175.  
PR 18-AUG-1999; 9905-0149426.  
PR 20-AUG-1999; 9905-0149722.  
PR 20-AUG-1999; 9905-0149723.  
PR 20-AUG-1999; 9905-0149929.  
PR 23-AUG-1999; 9905-0149902.  
PR 23-AUG-1999; 9905-0149930.  
PR 23-AUG-1999; 9905-0150566.  
PR 26-AUG-1999; 9905-0150884.  
PR 27-AUG-1999; 9905-0151065.  
PR 27-AUG-1999; 9905-0151066.  
PR 27-AUG-1999; 9905-0151080.  
PR 30-AUG-1999; 9905-0151303.  
PR 31-AUG-1999; 9905-0151438.  
PR 01-SEP-1999; 9905-0151930.  
PR 07-SEP-1999; 9905-0152363.  
PR 10-SEP-1999; 9905-0153070.  
PR 13-SEP-1999; 9905-0153758.  
PR 15-SEP-1999; 9905-0154018.  
PR 16-SEP-1999; 9905-0154039.

PR 20-SEP-1999; 9905-0154779.  
PR 22-SEP-1999; 9905-0155139.  
PR 23-SEP-1999; 9905-0155486.  
PR 24-SEP-1999; 9905-0155659.  
PR 28-SEP-1999; 9905-0156458.  
PR 29-SEP-1999; 9905-0156596.  
PR 04-OCT-1999; 9905-0157117.  
PR 05-OCT-1999; 9905-0157753.  
PR 06-OCT-1999; 9905-0157865.  
PR 07-OCT-1999; 9905-0158029.  
PR 08-OCT-1999; 9905-0158232.  
PR 12-OCT-1999; 9905-0158369.  
PR 13-OCT-1999; 9905-0159293.  
PR 13-OCT-1999; 9905-0159294.  
PR 13-OCT-1999; 9905-0159295.  
PR 14-OCT-1999; 9905-0159329.  
PR 14-OCT-1999; 9905-0159330.  
PR 14-OCT-1999; 9905-0159331.  
PR 14-OCT-1999; 9905-0159637.  
PR 14-OCT-1999; 9905-0159638.  
PR 18-OCT-1999; 9905-0159584.  
PR 21-OCT-1999; 9905-0160741.  
PR 21-OCT-1999; 9905-0160767.  
PR 21-OCT-1999; 9905-0160768.  
PR 21-OCT-1999; 9905-0160770.  
PR 21-OCT-1999; 9905-0160814.  
PR 21-OCT-1999; 9905-0160815.  
PR 22-OCT-1999; 9905-0160980.  
PR 22-OCT-1999; 9905-0160981.  
PR 22-OCT-1999; 9905-0160989.  
PR 25-OCT-1999; 9905-0161404.  
PR 25-OCT-1999; 9905-0161405.  
PR 26-OCT-1999; 9905-0161406.  
PR 26-OCT-1999; 9905-0161359.  
PR 26-OCT-1999; 9905-0161360.  
PR 26-OCT-1999; 9905-0161361.  
PR 28-OCT-1999; 9905-0161920.  
PR 28-OCT-1999; 9905-0161992.  
PR 28-OCT-1999; 9905-0161993.  
PR 29-OCT-1999; 9905-0162142.

Query Match 2.5%; Score 10; DB 21; Length 235;

Best Local Similarity 100.0%; Pred. No. 0.52; Mismatches 0; Indels 0; Gaps 0;

Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 48 pppppppppn 57

Db 79 pppppppppn 88

RESULT 5

AAG37672

ID AAG37672 standard; Protein; 235 AA.

XX AC AAG37672;

XX DT 18-OCT-2000 (first entry)

XX DE Arabidopsis thaliana protein fragment SEQ ID NO: 46360.

XX KW Protein identification; signal transduction pathway; metabolic pathway;

XX KW hybridisation assay; genetic mapping; gene expression control; promoter;

XX OS Arabidopsis thaliana.

XX PN EP1033405-A2.

XX PD 06-SEP-2000.

XX PF 25-FEB-2000; 2000EP-0301439.

XX PR 25-FEB-1999; 9905-0121825.



PR 05-MAR-1999; 99US-0123180.  
PR 09-MAR-1999; 99US-0123548.  
PR 23-MAR-1999; 99US-0125788.  
PR 25-MAR-1999; 99US-0126264.  
PR 29-MAR-1999; 99US-0126785.  
PR 01-APR-1999; 99US-0127462.  
PR 06-APR-1999; 99US-0128234.  
PR 08-APR-1999; 99US-0128714.  
PR 16-APR-1999; 99US-0129845.  
PR 19-APR-1999; 99US-0130077.  
PR 21-APR-1999; 99US-0130449.  
PR 23-APR-1999; 99US-0130510.  
PR 23-APR-1999; 99US-0130891.  
PR 28-APR-1999; 99US-0131449.  
PR 30-APR-1999; 99US-0132048.  
PR 30-APR-1999; 99US-0132407.  
PR 04-MAY-1999; 99US-0132484.  
PR 05-MAY-1999; 99US-0132485.  
PR 06-MAY-1999; 99US-0132486.  
PR 06-MAY-1999; 99US-0132487.  
PR 07-MAY-1999; 99US-0132863.  
PR 11-MAY-1999; 99US-0134256.  
PR 14-MAY-1999; 99US-0134218.  
PR 14-MAY-1999; 99US-0134219.  
PR 14-MAY-1999; 99US-0134221.  
PR 14-MAY-1999; 99US-0134370.  
PR 18-MAY-1999; 99US-0134768.  
PR 19-MAY-1999; 99US-0134941.  
PR 20-MAY-1999; 99US-0135124.  
PR 21-MAY-1999; 99US-0135353.  
PR 24-MAY-1999; 99US-0135629.  
PR 25-MAY-1999; 99US-0136021.  
PR 27-MAY-1999; 99US-0136392.  
PR 28-MAY-1999; 99US-0136782.  
PR 01-JUN-1999; 99US-0137222.  
PR 03-JUN-1999; 99US-0137528.  
PR 04-JUN-1999; 99US-0137502.  
PR 07-JUN-1999; 99US-0137724.  
PR 08-JUN-1999; 99US-0138094.  
PR 10-JUN-1999; 99US-0138540.  
PR 10-JUN-1999; 99US-0138847.  
PR 14-JUN-1999; 99US-0139119.  
PR 16-JUN-1999; 99US-0139452.  
PR 16-JUN-1999; 99US-0139453.  
PR 17-JUN-1999; 99US-0139492.  
PR 18-JUN-1999; 99US-0139454.  
PR 18-JUN-1999; 99US-0139455.  
PR 18-JUN-1999; 99US-0139456.  
PR 18-JUN-1999; 99US-0139457.  
PR 18-JUN-1999; 99US-0139458.  
PR 18-JUN-1999; 99US-0139459.  
PR 18-JUN-1999; 99US-0139460.  
PR 18-JUN-1999; 99US-0139461.  
PR 18-JUN-1999; 99US-0139462.  
PR 18-JUN-1999; 99US-0139463.  
PR 18-JUN-1999; 99US-0139750.  
PR 18-JUN-1999; 99US-0139763.  
PR 21-JUN-1999; 99US-0139817.  
PR 22-JUN-1999; 99US-0139899.  
PR 23-JUN-1999; 99US-0140353.  
PR 23-JUN-1999; 99US-0140354.  
PR 24-JUN-1999; 99US-0140695.  
PR 24-JUN-1999; 99US-0140823.  
PR 29-JUN-1999; 99US-0140991.  
PR 30-JUN-1999; 99US-0141287.  
PR 01-JUL-1999; 99US-0141842.  
PR 01-JUL-1999; 99US-0142154.  
PR 02-JUL-1999; 99US-0142055.  
PR 06-JUL-1999; 99US-0142390.  
PR 08-JUL-1999; 99US-0142803.  
PR 09-JUL-1999; 99US-0142920.  
PR 12-JUL-1999; 99US-0142977.  
PR 13-JUL-1999; 99US-0143542.

PR 14-JUL-1999; 99US-0143624.  
PR 15-JUL-1999; 99US-0144005.  
PR 16-JUL-1999; 99US-0144085.  
PR 16-JUL-1999; 99US-0144086.  
PR 19-JUL-1999; 99US-0144325.  
PR 19-JUL-1999; 99US-0144331.  
PR 19-JUL-1999; 99US-0144332.  
PR 19-JUL-1999; 99US-0144333.  
PR 19-JUL-1999; 99US-0144334.  
PR 19-JUL-1999; 99US-0144335.  
PR 20-JUL-1999; 99US-0144352.  
PR 20-JUL-1999; 99US-0144632.  
PR 20-JUL-1999; 99US-0144684.  
PR 21-JUL-1999; 99US-0144814.  
PR 21-JUL-1999; 99US-0145086.  
PR 21-JUL-1999; 99US-0145088.  
PR 22-JUL-1999; 99US-0145085.  
PR 22-JUL-1999; 99US-0145087.  
PR 22-JUL-1999; 99US-0145089.  
PR 22-JUL-1999; 99US-0145192.  
PR 23-JUL-1999; 99US-0145145.  
PR 23-JUL-1999; 99US-0145218.  
PR 23-JUL-1999; 99US-0145224.  
PR 26-JUL-1999; 99US-0145276.  
PR 27-JUL-1999; 99US-0145913.  
PR 27-JUL-1999; 99US-0145918.  
PR 27-JUL-1999; 99US-0145919.  
PR 28-JUL-1999; 99US-0145951.  
PR 02-AUG-1999; 99US-0146386.  
PR 02-AUG-1999; 99US-0146388.  
PR 03-AUG-1999; 99US-0146389.  
PR 03-AUG-1999; 99US-0147038.  
PR 04-AUG-1999; 99US-0147204.  
PR 04-AUG-1999; 99US-0147302.  
PR 05-AUG-1999; 99US-0147192.  
PR 05-AUG-1999; 99US-0147260.  
PR 06-AUG-1999; 99US-0147303.  
PR 06-AUG-1999; 99US-0147416.  
PR 09-AUG-1999; 99US-0147493.  
PR 09-AUG-1999; 99US-0147493.  
PR 10-AUG-1999; 99US-0147935.  
PR 11-AUG-1999; 99US-0148171.  
PR 11-AUG-1999; 99US-0148319.  
PR 12-AUG-1999; 99US-0148341.  
PR 13-AUG-1999; 99US-0148565.  
PR 13-AUG-1999; 99US-0148684.  
PR 16-AUG-1999; 99US-0149368.  
PR 17-AUG-1999; 99US-0149175.  
PR 18-AUG-1999; 99US-0149426.  
PR 20-AUG-1999; 99US-0149722.  
PR 20-AUG-1999; 99US-0149723.  
PR 20-AUG-1999; 99US-0149929.  
PR 23-AUG-1999; 99US-0149902.  
PR 23-AUG-1999; 99US-0149930.  
PR 25-AUG-1999; 99US-0150566.  
PR 26-AUG-1999; 99US-0150884.  
PR 27-AUG-1999; 99US-0151065.  
PR 27-AUG-1999; 99US-0151066.  
PR 27-AUG-1999; 99US-0151080.  
PR 30-AUG-1999; 99US-0151303.  
PR 31-AUG-1999; 99US-0151438.  
PR 01-SEP-1999; 99US-0151930.  
PR 07-SEP-1999; 99US-0152363.  
PR 10-SEP-1999; 99US-0153070.  
PR 13-SEP-1999; 99US-0153758.  
PR 15-SEP-1999; 99US-0154018.  
PR 16-SEP-1999; 99US-0154039.  
PR 20-SEP-1999; 99US-0154779.  
PR 22-SEP-1999; 99US-0155139.  
PR 23-SEP-1999; 99US-0155486.  
PR 24-SEP-1999; 99US-0155659.  
PR 28-SEP-1999; 99US-0156458.  
PR 29-SEP-1999; 99US-0156596.  
PR 04-OCT-1999; 99US-0157117.

PR	05-OCT-1999;	99US-0157753.	PR	08-APR-1999;	99US-0128714.
PR	06-OCT-1999;	99US-0157865.	PR	16-APR-1999;	99US-0129845.
PR	07-OCT-1999;	99US-0158029.	PR	19-APR-1999;	99US-0130077.
PR	08-OCT-1999;	99US-0158232.	PR	21-APR-1999;	99US-0130449.
PR	12-OCT-1999;	99US-0158369.	PR	23-APR-1999;	99US-0130510.
PR	13-OCT-1999;	99US-0159293.	PR	23-APR-1999;	99US-0130891.
PR	13-OCT-1999;	99US-0159294.	PR	28-APR-1999;	99US-0131449.
PR	13-OCT-1999;	99US-0159295.	PR	30-APR-1999;	99US-0132048.
PR	14-OCT-1999;	99US-0159329.	PR	30-APR-1999;	99US-0132407.
PR	14-OCT-1999;	99US-0159330.	PR	04-MAY-1999;	99US-0132484.
PR	14-OCT-1999;	99US-0159331.	PR	05-MAY-1999;	99US-0132485.
PR	14-OCT-1999;	99US-0159637.	PR	06-MAY-1999;	99US-0132486.
PR	14-OCT-1999;	99US-0159638.	PR	06-MAY-1999;	99US-0132487.
PR	18-OCT-1999;	99US-0159584.	PR	07-MAY-1999;	99US-0132863.
PR	21-OCT-1999;	99US-0160741.	PR	11-MAY-1999;	99US-0134256.
PR	21-OCT-1999;	99US-0160767.	PR	14-MAY-1999;	99US-0134218.
PR	21-OCT-1999;	99US-0160768.	PR	14-MAY-1999;	99US-0134219.
PR	21-OCT-1999;	99US-0160770.	PR	14-MAY-1999;	99US-0134221.
PR	21-OCT-1999;	99US-0160814.	PR	14-MAY-1999;	99US-0134370.
PR	21-OCT-1999;	99US-0160815.	PR	18-MAY-1999;	99US-0134768.
PR	22-OCT-1999;	99US-0160980.	PR	19-MAY-1999;	99US-0134941.
PR	22-OCT-1999;	99US-0160981.	PR	20-MAY-1999;	99US-0135124.
PR	22-OCT-1999;	99US-0160989.	PR	21-MAY-1999;	99US-0135353.
PR	25-OCT-1999;	99US-0161404.	PR	24-MAY-1999;	99US-0135629.
PR	25-OCT-1999;	99US-0161405.	PR	25-MAY-1999;	99US-0136021.
PR	25-OCT-1999;	99US-0161406.	PR	27-MAY-1999;	99US-0136392.
PR	26-OCT-1999;	99US-0161359.	PR	28-MAY-1999;	99US-0136782.
PR	26-OCT-1999;	99US-0161360.	PR	01-JUN-1999;	99US-0137222.
PR	26-OCT-1999;	99US-0161361.	PR	03-JUN-1999;	99US-0137528.
PR	28-OCT-1999;	99US-0161920.	PR	04-JUN-1999;	99US-0137502.
PR	28-OCT-1999;	99US-0161992.	PR	07-JUN-1999;	99US-0137724.
PR	28-OCT-1999;	99US-0161993.	PR	08-JUN-1999;	99US-0138094.
PR	29-OCT-1999;	99US-0162142.	PR	10-JUN-1999;	99US-0138540.
PR	29-OCT-1999;	99US-0162142.	PR	10-JUN-1999;	99US-0138847.
PR	29-OCT-1999;	99US-0162142.	PR	14-JUN-1999;	99US-0139119.
PR	29-OCT-1999;	99US-0162142.	PR	16-JUN-1999;	99US-0139452.
PR	29-OCT-1999;	99US-0162142.	PR	16-JUN-1999;	99US-0139453.
PR	29-OCT-1999;	99US-0162142.	PR	17-JUN-1999;	99US-0139492.
PR	29-OCT-1999;	99US-0162142.	PR	18-JUN-1999;	99US-0139454.
PR	29-OCT-1999;	99US-0162142.	PR	18-JUN-1999;	99US-0139455.
PR	29-OCT-1999;	99US-0162142.	PR	18-JUN-1999;	99US-0139456.
PR	29-OCT-1999;	99US-0162142.	PR	18-JUN-1999;	99US-0139457.
PR	29-OCT-1999;	99US-0162142.	PR	18-JUN-1999;	99US-0139458.
PR	29-OCT-1999;	99US-0162142.	PR	18-JUN-1999;	99US-0139459.
PR	29-OCT-1999;	99US-0162142.	PR	18-JUN-1999;	99US-0139460.
PR	29-OCT-1999;	99US-0162142.	PR	18-JUN-1999;	99US-0139461.
PR	29-OCT-1999;	99US-0162142.	PR	18-JUN-1999;	99US-0139462.
PR	29-OCT-1999;	99US-0162142.	PR	18-JUN-1999;	99US-0139463.
PR	29-OCT-1999;	99US-0162142.	PR	18-JUN-1999;	99US-0139750.
PR	29-OCT-1999;	99US-0162142.	PR	18-JUN-1999;	99US-0139763.
PR	29-OCT-1999;	99US-0162142.	PR	21-JUN-1999;	99US-0139817.
PR	29-OCT-1999;	99US-0162142.	PR	22-JUN-1999;	99US-0139899.
PR	29-OCT-1999;	99US-0162142.	PR	23-JUN-1999;	99US-0140353.
PR	29-OCT-1999;	99US-0162142.	PR	23-JUN-1999;	99US-0140354.
PR	29-OCT-1999;	99US-0162142.	PR	24-JUN-1999;	99US-0140695.
PR	29-OCT-1999;	99US-0162142.	PR	28-JUN-1999;	99US-0140823.
PR	29-OCT-1999;	99US-0162142.	PR	29-JUN-1999;	99US-0140991.
PR	29-OCT-1999;	99US-0162142.	PR	30-JUN-1999;	99US-0141287.
PR	29-OCT-1999;	99US-0162142.	PR	01-JUL-1999;	99US-0141842.
PR	29-OCT-1999;	99US-0162142.	PR	01-JUL-1999;	99US-0142154.
PR	29-OCT-1999;	99US-0162142.	PR	02-JUL-1999;	99US-0142055.
PR	29-OCT-1999;	99US-0162142.	PR	06-JUL-1999;	99US-0142390.
PR	29-OCT-1999;	99US-0162142.	PR	08-JUL-1999;	99US-0142803.
PR	29-OCT-1999;	99US-0162142.	PR	09-JUL-1999;	99US-0142920.
PR	29-OCT-1999;	99US-0162142.	PR	12-JUL-1999;	99US-0142977.
PR	29-OCT-1999;	99US-0162142.	PR	13-JUL-1999;	99US-0143542.
PR	29-OCT-1999;	99US-0162142.	PR	14-JUL-1999;	99US-0143624.
PR	29-OCT-1999;	99US-0162142.	PR	15-JUL-1999;	99US-0144005.
PR	29-OCT-1999;	99US-0162142.	PR	16-JUL-1999;	99US-0144085.
PR	29-OCT-1999;	99US-0162142.	PR	16-JUL-1999;	99US-0144086.
PR	29-OCT-1999;	99US-0162142.	PR	19-JUL-1999;	99US-0144325.
PR	29-OCT-1999;	99US-0162142.	PR	19-JUL-1999;	99US-0144331.
PR	29-OCT-1999;	99US-0162142.	PR	19-JUL-1999;	99US-0144332.

Query Match 2.5%; Score 10; DB 21; Length 235;

Best Local Similarity 100.0%; Pred. No. 0.52;

Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 48 PPPPPPPPPN 57

Db 79 PPPPPPPPPN 88

RESULT 6

AGL11271

ID AGL11271 standard; Protein; 237 AA.

XX AGL11271;

AC AGL11271;

XX 17-OCT-2000 (first entry)

XX Arabidopsis thaliana protein fragment SEQ ID NO: 9925.

DE Protein identification; signal transduction pathway; metabolic pathway;

KW hybridisation assay; genetic mapping; gene expression control; promoter;

KW termination sequence.

XX Arabidopsis thaliana.

OS Arabidopsis thaliana.

XX EP1033405-A2.

XX 06-SEP-2000.

XX 25-FEB-2000; 2000EP-0301439.

XX 25-FEB-1999; 99US-0121825.

PR 05-MAR-1999; 99US-0123180.

PR 09-MAR-1999; 99US-0123548.

PR 23-MAR-1999; 99US-0125788.

PR 25-MAR-1999; 99US-0126264.

PR 29-MAR-1999; 99US-0126785.

PR 01-APR-1999; 99US-0127462.

PR 06-APR-1999; 99US-0128234.

PR 19-JUL-1999;	99US-0144333.
PR 19-JUL-1999;	99US-0144334.
PR 19-JUL-1999;	99US-0144335.
PR 20-JUL-1999;	99US-0144352.
PR 20-JUL-1999;	99US-0144632.
PR 20-JUL-1999;	99US-0144884.
PR 21-JUL-1999;	99US-0144814.
PR 21-JUL-1999;	99US-0145086.
PR 21-JUL-1999;	99US-0145088.
PR 22-JUL-1999;	99US-0145085.
PR 22-JUL-1999;	99US-0145087.
PR 22-JUL-1999;	99US-0145089.
PR 22-JUL-1999;	99US-0145192.
PR 23-JUL-1999;	99US-0145145.
PR 23-JUL-1999;	99US-0145218.
PR 23-JUL-1999;	99US-0145224.
PR 26-JUL-1999;	99US-0145276.
PR 27-JUL-1999;	99US-0145913.
PR 27-JUL-1999;	99US-0145918.
PR 27-JUL-1999;	99US-0145919.
PR 28-JUL-1999;	99US-0145951.
PR 02-AUG-1999;	99US-0146386.
PR 02-AUG-1999;	99US-0146388.
PR 02-AUG-1999;	99US-0146389.
PR 03-AUG-1999;	99US-0147038.
PR 04-AUG-1999;	99US-0147204.
PR 04-AUG-1999;	99US-0147302.
PR 05-AUG-1999;	99US-0147192.
PR 05-AUG-1999;	99US-0147260.
PR 08-AUG-1999;	99US-0147303.
PR 06-AUG-1999;	99US-0147416.
PR 09-AUG-1999;	99US-0147493.
PR 09-AUG-1999;	99US-0147935.
PR 10-AUG-1999;	99US-0148171.
PR 11-AUG-1999;	99US-0148319.
PR 12-AUG-1999;	99US-0148341.
PR 13-AUG-1999;	99US-0148565.
PR 13-AUG-1999;	99US-0148684.
PR 16-AUG-1999;	99US-0149368.
PR 17-AUG-1999;	99US-0149175.
PR 18-AUG-1999;	99US-0149426.
PR 20-AUG-1999;	99US-0149722.
PR 20-AUG-1999;	99US-0149723.
PR 20-AUG-1999;	99US-0149929.
PR 23-AUG-1999;	99US-0149902.
PR 23-AUG-1999;	99US-0149930.
PR 25-AUG-1999;	99US-0150566.
PR 26-AUG-1999;	99US-0150884.
PR 27-AUG-1999;	99US-0151065.
PR 27-AUG-1999;	99US-0151066.
PR 27-AUG-1999;	99US-0151080.
PR 30-AUG-1999;	99US-0151303.
PR 31-AUG-1999;	99US-0151438.
PR 01-SEP-1999;	99US-0151930.
PR 07-SEP-1999;	99US-0152363.
PR 10-SEP-1999;	99US-0153070.
PR 13-SEP-1999;	99US-0153758.
PR 15-SEP-1999;	99US-0154018.
PR 16-SEP-1999;	99US-0154039.
PR 20-SEP-1999;	99US-0154779.
PR 22-SEP-1999;	99US-0155139.
PR 23-SEP-1999;	99US-0155486.
PR 24-SEP-1999;	99US-0155659.
PR 28-SEP-1999;	99US-0156458.
PR 29-SEP-1999;	99US-0156596.
PR 04-OCT-1999;	99US-0157117.
PR 05-OCT-1999;	99US-0157753.
PR 05-OCT-1999;	99US-0157865.
PR 07-OCT-1999;	99US-0158029.
PR 08-OCT-1999;	99US-0158232.
PR 12-OCT-1999;	99US-0158369.
PR 13-OCT-1999;	99US-0159293.
PR 13-OCT-1999;	99US-0159294.
PR 13-OCT-1999;	99US-0159295.
PR 14-OCT-1999;	99US-0159329.
PR 14-OCT-1999;	99US-0159330.
PR 14-OCT-1999;	99US-0159331.
PR 14-OCT-1999;	99US-0159637.
PR 14-OCT-1999;	99US-0159638.
PR 18-OCT-1999;	99US-0159584.
PR 21-OCT-1999;	99US-0160741.
PR 21-OCT-1999;	99US-0160767.
PR 21-OCT-1999;	99US-0160768.
PR 21-OCT-1999;	99US-0160770.
PR 21-OCT-1999;	99US-0160814.
PR 21-OCT-1999;	99US-0160815.
PR 22-OCT-1999;	99US-0160980.
PR 22-OCT-1999;	99US-0160981.
PR 22-OCT-1999;	99US-0160989.
PR 25-OCT-1999;	99US-0161404.
PR 25-OCT-1999;	99US-0161405.
PR 25-OCT-1999;	99US-0161406.
PR 26-OCT-1999;	99US-0161359.
PR 26-OCT-1999;	99US-0161360.
PR 26-OCT-1999;	99US-0161361.
PR 28-OCT-1999;	99US-0161820.
PR 28-OCT-1999;	99US-0161992.
PR 28-OCT-1999;	99US-0161993.
PR 29-OCT-1999;	99US-0162142.
Query Match 2.5%; Score 10; DB 21; Length 237;	
Best Local Similarity 100.0%; Pred. No. 0.52;	
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;	
QY 48 pppppppppn 57	
Db 81 pppppppppn 90	
RESULT 7	
AAG37671	
ID AAG37671 standard; Protein; 237 AA.	
XX	
AC AAG37671;	
XX	
DT 18-OCT-2000 (first entry)	
XX	
DE Arabidopsis thaliana protein fragment SEQ ID NO: 46359.	
XX	
KW Protein identification; signal transduction pathway; metabolic pathway;	
KW hybridisation assay; genetic mapping; gene expression control; promoter;	
KW termination sequence.	
XX	
OS Arabidopsis thaliana.	
XX	
PN EP1033405-A2.	
XX	
PD 06-SEP-2000.	
XX	
PF 25-FEB-2000; 2000EP-0301439.	
XX	
PR 25-FEB-1999;	99US-0121825.
PR 05-MAR-1999;	99US-0123180.
PR 09-MAR-1999;	99US-0123548.
PR 23-MAR-1999;	99US-0125788.
PR 25-MAR-1999;	99US-0126264.
PR 29-MAR-1999;	99US-0126785.
PR 01-APR-1999;	99US-0127462.
PR 06-APR-1999;	99US-0128234.
PR 08-APR-1999;	99US-0128714.
PR 16-APR-1999;	99US-0129845.
PR 19-APR-1999;	99US-0130077.
PR 21-APR-1999;	99US-0130449.
PR 23-APR-1999;	99US-0130510.
PR 23-APR-1999;	99US-0130891.
PR 28-APR-1999;	99US-0131449.

PR 30-APR-1999; 99US-0132048.  
PR 30-APR-1999; 99US-0132407.  
PR 04-MAY-1999; 99US-0132484.  
PR 05-MAY-1999; 99US-0132485.  
PR 06-MAY-1999; 99US-0132486.  
PR 06-MAY-1999; 99US-0132487.  
PR 07-MAY-1999; 99US-0132863.  
PR 11-MAY-1999; 99US-0134256.  
PR 14-MAY-1999; 99US-0134218.  
PR 14-MAY-1999; 99US-0134219.  
PR 14-MAY-1999; 99US-0134221.  
PR 14-MAY-1999; 99US-0134370.  
PR 18-MAY-1999; 99US-0134768.  
PR 19-MAY-1999; 99US-0134941.  
PR 20-MAY-1999; 99US-0135124.  
PR 21-MAY-1999; 99US-0135353.  
PR 24-MAY-1999; 99US-0135629.  
PR 25-MAY-1999; 99US-0136021.  
PR 27-MAY-1999; 99US-0136392.  
PR 28-MAY-1999; 99US-0136782.  
PR 01-JUN-1999; 99US-0137222.  
PR 03-JUN-1999; 99US-0137528.  
PR 04-JUN-1999; 99US-0137502.  
PR 07-JUN-1999; 99US-0137724.  
PR 08-JUN-1999; 99US-0138094.  
PR 10-JUN-1999; 99US-0138540.  
PR 10-JUN-1999; 99US-0138847.  
PR 14-JUN-1999; 99US-0139119.  
PR 16-JUN-1999; 99US-0139452.  
PR 16-JUN-1999; 99US-0139453.  
PR 17-JUN-1999; 99US-0139492.  
PR 18-JUN-1999; 99US-0139454.  
PR 18-JUN-1999; 99US-0139455.  
PR 18-JUN-1999; 99US-0139456.  
PR 18-JUN-1999; 99US-0139457.  
PR 18-JUN-1999; 99US-0139458.  
PR 18-JUN-1999; 99US-0139459.  
PR 18-JUN-1999; 99US-0139460.  
PR 18-JUN-1999; 99US-0139461.  
PR 18-JUN-1999; 99US-0139462.  
PR 18-JUN-1999; 99US-0139463.  
PR 18-JUN-1999; 99US-0139750.  
PR 18-JUN-1999; 99US-0139763.  
PR 21-JUN-1999; 99US-0139817.  
PR 22-JUN-1999; 99US-0139899.  
PR 23-JUN-1999; 99US-0140353.  
PR 23-JUN-1999; 99US-0140354.  
PR 24-JUN-1999; 99US-0140695.  
PR 28-JUN-1999; 99US-0140823.  
PR 29-JUN-1999; 99US-0140991.  
PR 30-JUN-1999; 99US-0141287.  
PR 01-JUL-1999; 99US-0141842.  
PR 01-JUL-1999; 99US-0142154.  
PR 02-JUL-1999; 99US-0142055.  
PR 06-JUL-1999; 99US-0142390.  
PR 08-JUL-1999; 99US-0142803.  
PR 09-JUL-1999; 99US-0142920.  
PR 12-JUL-1999; 99US-0142977.  
PR 13-JUL-1999; 99US-0143542.  
PR 14-JUL-1999; 99US-0143624.  
PR 15-JUL-1999; 99US-0144005.  
PR 16-JUL-1999; 99US-0144085.  
PR 16-JUL-1999; 99US-0144086.  
PR 19-JUL-1999; 99US-0144325.  
PR 19-JUL-1999; 99US-0144331.  
PR 19-JUL-1999; 99US-0144332.  
PR 19-JUL-1999; 99US-0144333.  
PR 19-JUL-1999; 99US-0144334.  
PR 19-JUL-1999; 99US-0144335.  
PR 20-JUL-1999; 99US-0144352.  
PR 20-JUL-1999; 99US-0144632.  
PR 20-JUL-1999; 99US-0144884.  
PR 21-JUL-1999; 99US-0144814.

PR 21-JUL-1999; 99US-0145086.  
PR 21-JUL-1999; 99US-0145088.  
PR 22-JUL-1999; 99US-0145085.  
PR 22-JUL-1999; 99US-0145087.  
PR 22-JUL-1999; 99US-0145089.  
PR 22-JUL-1999; 99US-0145192.  
PR 23-JUL-1999; 99US-0145145.  
PR 23-JUL-1999; 99US-0145218.  
PR 23-JUL-1999; 99US-0145224.  
PR 26-JUL-1999; 99US-0145276.  
PR 27-JUL-1999; 99US-0145913.  
PR 27-JUL-1999; 99US-0145918.  
PR 27-JUL-1999; 99US-0145919.  
PR 28-JUL-1999; 99US-0145951.  
PR 02-AUG-1999; 99US-0146386.  
PR 02-AUG-1999; 99US-0146388.  
PR 02-AUG-1999; 99US-0146389.  
PR 03-AUG-1999; 99US-0147038.  
PR 04-AUG-1999; 99US-0147204.  
PR 04-AUG-1999; 99US-0147302.  
PR 05-AUG-1999; 99US-0147192.  
PR 05-AUG-1999; 99US-0147260.  
PR 06-AUG-1999; 99US-0147303.  
PR 06-AUG-1999; 99US-0147416.  
PR 09-AUG-1999; 99US-0147493.  
PR 09-AUG-1999; 99US-0147935.  
PR 10-AUG-1999; 99US-0148171.  
PR 11-AUG-1999; 99US-0148319.  
PR 12-AUG-1999; 99US-0148341.  
PR 13-AUG-1999; 99US-0148565.  
PR 13-AUG-1999; 99US-0148684.  
PR 16-AUG-1999; 99US-0149368.  
PR 17-AUG-1999; 99US-0149175.  
PR 18-AUG-1999; 99US-0149426.  
PR 20-AUG-1999; 99US-0149722.  
PR 20-AUG-1999; 99US-0149723.  
PR 20-AUG-1999; 99US-0149929.  
PR 23-AUG-1999; 99US-0149902.  
PR 23-AUG-1999; 99US-0149930.  
PR 25-AUG-1999; 99US-0150566.  
PR 26-AUG-1999; 99US-0150884.  
PR 27-AUG-1999; 99US-0151065.  
PR 27-AUG-1999; 99US-0151066.  
PR 27-AUG-1999; 99US-0151080.  
PR 30-AUG-1999; 99US-0151303.  
PR 31-AUG-1999; 99US-0151438.  
PR 01-SEP-1999; 99US-0151930.  
PR 07-SEP-1999; 99US-0152363.  
PR 10-SEP-1999; 99US-0153070.  
PR 13-SEP-1999; 99US-0153758.  
PR 15-SEP-1999; 99US-0154018.  
PR 16-SEP-1999; 99US-0154039.  
PR 20-SEP-1999; 99US-0154779.  
PR 22-SEP-1999; 99US-0155139.  
PR 23-SEP-1999; 99US-0155486.  
PR 24-SEP-1999; 99US-0155659.  
PR 28-SEP-1999; 99US-0156458.  
PR 29-SEP-1999; 99US-0156596.  
PR 04-OCT-1999; 99US-0157117.  
PR 05-OCT-1999; 99US-0157753.  
PR 06-OCT-1999; 99US-0157865.  
PR 07-OCT-1999; 99US-0158029.  
PR 08-OCT-1999; 99US-0158232.  
PR 12-OCT-1999; 99US-0158369.  
PR 13-OCT-1999; 99US-0159293.  
PR 13-OCT-1999; 99US-0159294.  
PR 13-OCT-1999; 99US-0159295.  
PR 14-OCT-1999; 99US-0159329.  
PR 14-OCT-1999; 99US-0159330.  
PR 14-OCT-1999; 99US-0159331.  
PR 14-OCT-1999; 99US-0159637.  
PR 14-OCT-1999; 99US-0159638.  
PR 18-OCT-1999; 99US-0159584.

PR	21-OCT-1999;	99US-0160741.	PR	11-MAY-1999;	99US-0134256.
PR	21-OCT-1999;	99US-0160767.	PR	14-MAY-1999;	99US-0134218.
PR	21-OCT-1999;	99US-0160768.	PR	14-MAY-1999;	99US-0134219.
PR	21-OCT-1999;	99US-0160770.	PR	14-MAY-1999;	99US-0134221.
PR	21-OCT-1999;	99US-0160814.	PR	14-MAY-1999;	99US-0134370.
PR	21-OCT-1999;	99US-0160815.	PR	18-MAY-1999;	99US-0134768.
PR	22-OCT-1999;	99US-0160980.	PR	19-MAY-1999;	99US-0134941.
PR	22-OCT-1999;	99US-0160981.	PR	20-MAY-1999;	99US-0135124.
PR	22-OCT-1999;	99US-0160989.	PR	21-MAY-1999;	99US-0135353.
PR	25-OCT-1999;	99US-0161404.	PR	21-MAY-1999;	99US-0135629.
PR	25-OCT-1999;	99US-0161405.	PR	24-MAY-1999;	99US-0135629.
PR	25-OCT-1999;	99US-0161406.	PR	25-MAY-1999;	99US-0136021.
PR	26-OCT-1999;	99US-0161359.	PR	27-MAY-1999;	99US-0136392.
PR	26-OCT-1999;	99US-0161360.	PR	28-MAY-1999;	99US-0136782.
PR	26-OCT-1999;	99US-0161361.	PR	01-JUN-1999;	99US-0137222.
PR	28-OCT-1999;	99US-0161920.	PR	03-JUN-1999;	99US-0137528.
PR	28-OCT-1999;	99US-0161922.	PR	04-JUN-1999;	99US-0137502.
PR	28-OCT-1999;	99US-0161993.	PR	07-JUN-1999;	99US-0137724.
PR	29-OCT-1999;	99US-0162142.	PR	08-JUN-1999;	99US-0138094.
PR	29-OCT-1999;	99US-0162142.	PR	10-JUN-1999;	99US-0138540.
PR	29-OCT-1999;	99US-0162142.	PR	10-JUN-1999;	99US-0138847.
PR	29-OCT-1999;	99US-0162142.	PR	14-JUN-1999;	99US-0139119.
PR	29-OCT-1999;	99US-0162142.	PR	16-JUN-1999;	99US-0139452.
PR	29-OCT-1999;	99US-0162142.	PR	16-JUN-1999;	99US-0139453.
PR	29-OCT-1999;	99US-0162142.	PR	17-JUN-1999;	99US-0139492.
PR	29-OCT-1999;	99US-0162142.	PR	18-JUN-1999;	99US-0139454.
PR	29-OCT-1999;	99US-0162142.	PR	18-JUN-1999;	99US-0139455.
PR	29-OCT-1999;	99US-0162142.	PR	18-JUN-1999;	99US-0139456.
PR	29-OCT-1999;	99US-0162142.	PR	18-JUN-1999;	99US-0139457.
PR	29-OCT-1999;	99US-0162142.	PR	18-JUN-1999;	99US-0139458.
PR	29-OCT-1999;	99US-0162142.	PR	18-JUN-1999;	99US-0139459.
PR	29-OCT-1999;	99US-0162142.	PR	18-JUN-1999;	99US-0139460.
PR	29-OCT-1999;	99US-0162142.	PR	18-JUN-1999;	99US-0139461.
PR	29-OCT-1999;	99US-0162142.	PR	18-JUN-1999;	99US-0139462.
PR	29-OCT-1999;	99US-0162142.	PR	18-JUN-1999;	99US-0139463.
PR	29-OCT-1999;	99US-0162142.	PR	18-JUN-1999;	99US-0139463.
PR	29-OCT-1999;	99US-0162142.	PR	18-JUN-1999;	99US-0139750.
PR	29-OCT-1999;	99US-0162142.	PR	18-JUN-1999;	99US-0139763.
PR	29-OCT-1999;	99US-0162142.	PR	21-JUN-1999;	99US-0139817.
PR	29-OCT-1999;	99US-0162142.	PR	22-JUN-1999;	99US-0139899.
PR	29-OCT-1999;	99US-0162142.	PR	23-JUN-1999;	99US-0140353.
PR	29-OCT-1999;	99US-0162142.	PR	23-JUN-1999;	99US-0140354.
PR	29-OCT-1999;	99US-0162142.	PR	24-JUN-1999;	99US-0140695.
PR	29-OCT-1999;	99US-0162142.	PR	28-JUN-1999;	99US-0140823.
PR	29-OCT-1999;	99US-0162142.	PR	29-JUN-1999;	99US-0140991.
PR	29-OCT-1999;	99US-0162142.	PR	30-JUN-1999;	99US-0141287.
PR	29-OCT-1999;	99US-0162142.	PR	01-JUL-1999;	99US-0141842.
PR	29-OCT-1999;	99US-0162142.	PR	01-JUL-1999;	99US-0142154.
PR	29-OCT-1999;	99US-0162142.	PR	02-JUL-1999;	99US-0142055.
PR	29-OCT-1999;	99US-0162142.	PR	06-JUL-1999;	99US-0142390.
PR	29-OCT-1999;	99US-0162142.	PR	08-JUL-1999;	99US-0142803.
PR	29-OCT-1999;	99US-0162142.	PR	09-JUL-1999;	99US-0142920.
PR	29-OCT-1999;	99US-0162142.	PR	12-JUL-1999;	99US-0142977.
PR	29-OCT-1999;	99US-0162142.	PR	13-JUL-1999;	99US-0143542.
PR	29-OCT-1999;	99US-0162142.	PR	14-JUL-1999;	99US-0143624.
PR	29-OCT-1999;	99US-0162142.	PR	15-JUL-1999;	99US-0144005.
PR	29-OCT-1999;	99US-0162142.	PR	16-JUL-1999;	99US-0144085.
PR	29-OCT-1999;	99US-0162142.	PR	16-JUL-1999;	99US-0144086.
PR	29-OCT-1999;	99US-0162142.	PR	19-JUL-1999;	99US-0144325.
PR	29-OCT-1999;	99US-0162142.	PR	19-JUL-1999;	99US-0144331.
PR	29-OCT-1999;	99US-0162142.	PR	19-JUL-1999;	99US-0144332.
PR	29-OCT-1999;	99US-0162142.	PR	19-JUL-1999;	99US-0144333.
PR	29-OCT-1999;	99US-0162142.	PR	19-JUL-1999;	99US-0144334.
PR	29-OCT-1999;	99US-0162142.	PR	19-JUL-1999;	99US-0144335.
PR	29-OCT-1999;	99US-0162142.	PR	20-JUL-1999;	99US-0144352.
PR	29-OCT-1999;	99US-0162142.	PR	20-JUL-1999;	99US-0144632.
PR	29-OCT-1999;	99US-0162142.	PR	20-JUL-1999;	99US-0144884.
PR	29-OCT-1999;	99US-0162142.	PR	21-JUL-1999;	99US-0144814.
PR	29-OCT-1999;	99US-0162142.	PR	21-JUL-1999;	99US-0145086.
PR	29-OCT-1999;	99US-0162142.	PR	21-JUL-1999;	99US-0145088.
PR	29-OCT-1999;	99US-0162142.	PR	22-JUL-1999;	99US-0145085.
PR	29-OCT-1999;	99US-0162142.	PR	22-JUL-1999;	99US-0145087.
PR	29-OCT-1999;	99US-0162142.	PR	22-JUL-1999;	99US-0145089.
PR	29-OCT-1999;	99US-0162142.	PR	22-JUL-1999;	99US-0145192.
PR	29-OCT-1999;	99US-0162142.	PR	23-JUL-1999;	99US-0145145.

Query Match 2.5%; Score 10; DB 21; Length 237;  
Best Local Similarity 100.0%; Pred. No. 0.52;  
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 48 PPPPPPPPN 57  
Dd 81 PPPPPPPPN 90

RESULT 8  
AAG39269  
ID AAG39269 standard; Protein; 237 AA.  
XX  
AC AAG39269;  
XX  
DT 18-OCT-2000 (first entry)  
XX  
DE Arabidopsis thaliana protein fragment SEQ ID NO: 48563.  
XX  
KW Protein identification; signal transduction pathway; metabolic pathway;  
KW hybridisation assay; genetic mapping; gene expression control; promoter;  
KW termination sequence.

OS Arabidopsis thaliana.  
XX  
PN EP1033405-A2.  
XX  
PD 06-SEP-2000.  
XX  
PF 25-FEB-2000; 2000EP-0301439.  
XX  
XX 25-FEB-1999; 99US-0121825.  
PR 05-MAR-1999; 99US-0123180.  
PR 09-MAR-1999; 99US-0123548.  
PR 23-MAR-1999; 99US-0125788.  
PR 25-MAR-1999; 99US-0126264.  
PR 29-MAR-1999; 99US-0126785.  
PR 01-APR-1999; 99US-0127462.  
PR 06-APR-1999; 99US-0128234.  
PR 08-APR-1999; 99US-0128714.  
PR 16-APR-1999; 99US-0129845.  
PR 19-APR-1999; 99US-0130077.  
PR 21-APR-1999; 99US-0130449.  
PR 23-APR-1999; 99US-0130510.  
PR 28-APR-1999; 99US-0130891.  
PR 28-APR-1999; 99US-0131449.  
PR 30-APR-1999; 99US-0132048.  
PR 30-APR-1999; 99US-0132407.  
PR 04-MAY-1999; 99US-0132484.  
PR 05-MAY-1999; 99US-0132485.  
PR 06-MAY-1999; 99US-0132486.  
PR 06-MAY-1999; 99US-0132487.  
PR 07-MAY-1999; 99US-0132863.

PR 23-JUL-1999; 99US-0145218.  
PR 26-JUL-1999; 99US-0145224.  
PR 26-JUL-1999; 99US-0145276.  
PR 27-JUL-1999; 99US-0145913.  
PR 27-JUL-1999; 99US-0145918.  
PR 27-JUL-1999; 99US-0145919.  
PR 28-JUL-1999; 99US-0145951.  
PR 02-AUG-1999; 99US-0146386.  
PR 02-AUG-1999; 99US-0146388.  
PR 02-AUG-1999; 99US-0146389.  
PR 03-AUG-1999; 99US-0147038.  
PR 04-AUG-1999; 99US-0147204.  
PR 04-AUG-1999; 99US-0147302.  
PR 05-AUG-1999; 99US-0147192.  
PR 05-AUG-1999; 99US-0147260.  
PR 06-AUG-1999; 99US-0147303.  
PR 06-AUG-1999; 99US-0147416.  
PR 09-AUG-1999; 99US-0147493.  
PR 09-AUG-1999; 99US-0147935.  
PR 10-AUG-1999; 99US-0148171.  
PR 11-AUG-1999; 99US-0148319.  
PR 12-AUG-1999; 99US-0148341.  
PR 13-AUG-1999; 99US-0148565.  
PR 13-AUG-1999; 99US-0148684.  
PR 16-AUG-1999; 99US-0149368.  
PR 17-AUG-1999; 99US-0149175.  
PR 18-AUG-1999; 99US-0149426.  
PR 20-AUG-1999; 99US-0149723.  
PR 20-AUG-1999; 99US-0149723.  
PR 20-AUG-1999; 99US-0149929.  
PR 23-AUG-1999; 99US-0149902.  
PR 23-AUG-1999; 99US-0149930.  
PR 25-AUG-1999; 99US-0150566.  
PR 26-AUG-1999; 99US-0150884.  
PR 27-AUG-1999; 99US-0151065.  
PR 27-AUG-1999; 99US-0151066.  
PR 27-AUG-1999; 99US-0151080.  
PR 30-AUG-1999; 99US-0151303.  
PR 31-AUG-1999; 99US-0151438.  
PR 01-SEP-1999; 99US-0151930.  
PR 07-SEP-1999; 99US-0152363.  
PR 10-SEP-1999; 99US-0153070.  
PR 13-SEP-1999; 99US-0153758.  
PR 15-SEP-1999; 99US-0154018.  
PR 16-SEP-1999; 99US-0154039.  
PR 20-SEP-1999; 99US-0154779.  
PR 22-SEP-1999; 99US-0155139.  
PR 23-SEP-1999; 99US-0155486.  
PR 24-SEP-1999; 99US-0155659.  
PR 28-SEP-1999; 99US-0156458.  
PR 29-SEP-1999; 99US-0156596.  
PR 04-OCT-1999; 99US-0157117.  
PR 05-OCT-1999; 99US-0157753.  
PR 06-OCT-1999; 99US-0157865.  
PR 07-OCT-1999; 99US-0158029.  
PR 08-OCT-1999; 99US-0158232.  
PR 12-OCT-1999; 99US-0158369.  
PR 13-OCT-1999; 99US-0159293.  
PR 13-OCT-1999; 99US-0159294.  
PR 13-OCT-1999; 99US-0159295.  
PR 14-OCT-1999; 99US-0159329.  
PR 14-OCT-1999; 99US-0159330.  
PR 14-OCT-1999; 99US-0159331.  
PR 14-OCT-1999; 99US-0159637.  
PR 14-OCT-1999; 99US-0159638.  
PR 18-OCT-1999; 99US-0159584.  
PR 21-OCT-1999; 99US-0160741.  
PR 21-OCT-1999; 99US-0160767.  
PR 21-OCT-1999; 99US-0160768.  
PR 21-OCT-1999; 99US-0160770.  
PR 21-OCT-1999; 99US-0160814.  
PR 21-OCT-1999; 99US-0160815.  
PR 22-OCT-1999; 99US-0160980.

PR 22-OCT-1999; 99US-0160981.  
PR 22-OCT-1999; 99US-0160989.  
PR 25-OCT-1999; 99US-0161404.  
PR 25-OCT-1999; 99US-0161405.  
PR 25-OCT-1999; 99US-0161406.  
PR 26-OCT-1999; 99US-0161359.  
PR 26-OCT-1999; 99US-0161360.  
PR 26-OCT-1999; 99US-0161361.  
PR 28-OCT-1999; 99US-0161920.  
PR 28-OCT-1999; 99US-0161992.  
PR 28-OCT-1999; 99US-0161993.  
PR 29-OCT-1999; 99US-0162142.  
  
Query Match 2.5%; Score 10; DB 21; Length 237;  
Best Local Similarity 100.0%; Pred. No. 0.52;  
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY 48 PPPPPPPPPN 57  
Db 81 PPPPPPPPPN 90  
  
RESULT 9  
AAG11270  
ID AAG11270 standard; Protein; 240 AA.  
XX  
AC AAG11270;  
XX  
DT 17-OCT-2000 (first entry)  
XX  
DE Arabidopsis thaliana protein fragment SEQ ID NO: 9924.  
XX  
KW Protein identification; signal transduction pathway; metabolic pathway;  
KW hybridisation assay; genetic mapping; gene expression control; promoter;  
KW termination sequence.  
XX  
OS Arabidopsis thaliana.  
XX  
PN EP1033405-A2.  
XX  
PD 06-SEP-2000.  
XX  
PF 25-FEB-2000; 2000EP-0301439.  
XX  
PR 25-FEB-1999; 99US-0121825.  
PR 05-MAR-1999; 99US-0123180.  
PR 09-MAR-1999; 99US-0123548.  
PR 23-MAR-1999; 99US-0123788.  
PR 29-MAR-1999; 99US-0126264.  
PR 01-APR-1999; 99US-0126785.  
PR 01-APR-1999; 99US-0127462.  
PR 06-APR-1999; 99US-0128234.  
PR 08-APR-1999; 99US-0128714.  
PR 16-APR-1999; 99US-0129845.  
PR 19-APR-1999; 99US-0130077.  
PR 21-APR-1999; 99US-0130449.  
PR 23-APR-1999; 99US-0130510.  
PR 23-APR-1999; 99US-0130891.  
PR 28-APR-1999; 99US-0131449.  
PR 30-APR-1999; 99US-0132048.  
PR 30-APR-1999; 99US-0132407.  
PR 04-MAY-1999; 99US-0132484.  
PR 05-MAY-1999; 99US-0132485.  
PR 06-MAY-1999; 99US-0132486.  
PR 06-MAY-1999; 99US-0132487.  
PR 07-MAY-1999; 99US-0132863.  
PR 11-MAY-1999; 99US-0134256.  
PR 14-MAY-1999; 99US-0134218.  
PR 14-MAY-1999; 99US-0134219.  
PR 14-MAY-1999; 99US-0134221.  
PR 14-MAY-1999; 99US-0134370.  
PR 18-MAY-1999; 99US-0134768.  
PR 19-MAY-1999; 99US-0134941.



```
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match      2.5%; Score 10; DB 21; Length 240;
Best Local Similarity 100.0%; Pred. No. 0.53;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 48 PPPPPPPPN 57
Db 84 ppppppppn 93

RESULT 10
AAG37670
ID AAG37670 standard; Protein; 240 AA.
XX
AC AAG37670;
XX
DT 18-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana protein fragment SEQ ID NO: 46358.
XX
KW Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
KW termination sequence.
XX
OS Arabidopsis thaliana.
XX
PN EP1033405-A2.
XX
PD 06-SEP-2000.
XX
PF 25-FEB-2000; 2000EP-0301439.
XX
PR 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 08-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 28-APR-1999; 99US-0130891.
PR 30-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.

PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 17-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
```



	Best Local Similarity	100.0%;	Pred. No.	0.53;	Matches	10;	Conservative	0;	Mismatches	0;	Indels	0;	Gaps	0;
QY	48	pppppppppn	57											
Db	84	pppppppppn	93											
RESULT ll														
ID	AAG39268													
ID	AAG39268	standard; Protein; 240 AA.												
AC	AAG39268;													
XX	18-OCT-2000	(first entry)												
DE	Arabidopsis thaliana	protein fragment SEQ ID NO: 48562.												
KW	Protein identification;	signal transduction pathway; metabolic pathway;												
KW	hybridisation assay;	genetic mapping; gene expression control; promoter;												
XX	termination sequence.													
OS	Arabidopsis thaliana.													
PN	EP1033405-A2.													
XX	06-SEP-2000.													
XX	25-FEB-2000;	2000EP-0301439.												
PR	25-FEB-1999;	99US-0121825.												
PR	05-MAR-1999;	99US-0123180.												
PR	09-MAR-1999;	99US-0123548.												
PR	23-MAR-1999;	99US-0125788.												
PR	25-MAR-1999;	99US-0126264.												
PR	29-MAR-1999;	99US-0126785.												
PR	01-APR-1999;	99US-0127462.												
PR	06-APR-1999;	99US-0128234.												
PR	08-APR-1999;	99US-0128714.												
PR	16-APR-1999;	99US-0129845.												
PR	19-APR-1999;	99US-0130077.												
PR	21-APR-1999;	99US-0130449.												
PR	23-APR-1999;	99US-0130510.												
PR	28-APR-1999;	99US-0130891.												
PR	30-APR-1999;	99US-0131449.												
PR	30-APR-1999;	99US-0132048.												
PR	04-MAY-1999;	99US-0132407.												
PR	05-MAY-1999;	99US-0132484.												
PR	06-MAY-1999;	99US-0132485.												
PR	07-MAY-1999;	99US-0132486.												
PR	11-MAY-1999;	99US-0132863.												
PR	14-MAY-1999;	99US-0134256.												
PR	14-MAY-1999;	99US-0134218.												
PR	14-MAY-1999;	99US-0134219.												
PR	14-MAY-1999;	99US-0134221.												
PR	18-MAY-1999;	99US-0134370.												
PR	19-MAY-1999;	99US-0134768.												
PR	20-MAY-1999;	99US-0135124.												
PR	21-MAY-1999;	99US-0135353.												
PR	24-MAY-1999;	99US-0135629.												
PR	25-MAY-1999;	99US-0136021.												
PR	27-MAY-1999;	99US-0136392.												
PR	28-MAY-1999;	99US-0136782.												
PR	01-JUN-1999;	99US-0137222.												
PR	03-JUN-1999;	99US-0137528.												
PR	04-JUN-1999;	99US-0137502.												
PR	07-JUN-1999;	99US-0137724.												
PR	08-JUN-1999;	99US-0138094.												
PR	10-JUN-1999;	99US-0138540.												
PR	10-JUN-1999;	99US-0138847.												
PR	14-JUN-1999;	99US-0139119.												

PR 16-JUN-1999; 99US-0139452.  
PR 16-JUN-1999; 99US-0139453.  
PR 17-JUN-1999; 99US-0139492.  
PR 18-JUN-1999; 99US-0139454.  
PR 18-JUN-1999; 99US-0139455.  
PR 18-JUN-1999; 99US-0139456.  
PR 18-JUN-1999; 99US-0139457.  
PR 18-JUN-1999; 99US-0139458.  
PR 18-JUN-1999; 99US-0139459.  
PR 18-JUN-1999; 99US-0139460.  
PR 18-JUN-1999; 99US-0139461.  
PR 18-JUN-1999; 99US-0139462.  
PR 18-JUN-1999; 99US-0139463.  
PR 18-JUN-1999; 99US-0139750.  
PR 18-JUN-1999; 99US-0139763.  
PR 21-JUN-1999; 99US-0139817.  
PR 22-JUN-1999; 99US-0139889.  
PR 23-JUN-1999; 99US-0140353.  
PR 23-JUN-1999; 99US-0140354.  
PR 24-JUN-1999; 99US-0140695.  
PR 28-JUN-1999; 99US-0140823.  
PR 29-JUN-1999; 99US-0140951.  
PR 30-JUN-1999; 99US-0141287.  
PR 01-JUL-1999; 99US-0141842.  
PR 01-JUL-1999; 99US-0142154.  
PR 02-JUL-1999; 99US-0142055.  
PR 06-JUL-1999; 99US-0142390.  
PR 08-JUL-1999; 99US-0142803.  
PR 09-JUL-1999; 99US-0142820.  
PR 12-JUL-1999; 99US-0142977.  
PR 13-JUL-1999; 99US-0143542.  
PR 14-JUL-1999; 99US-0143624.  
PR 15-JUL-1999; 99US-0144005.  
PR 16-JUL-1999; 99US-0144085.  
PR 16-JUL-1999; 99US-0144086.  
PR 19-JUL-1999; 99US-0144325.  
PR 19-JUL-1999; 99US-0144331.  
PR 19-JUL-1999; 99US-0144332.  
PR 19-JUL-1999; 99US-0144333.  
PR 19-JUL-1999; 99US-0144334.  
PR 19-JUL-1999; 99US-0144335.  
PR 20-JUL-1999; 99US-0144352.  
PR 20-JUL-1999; 99US-0144632.  
PR 20-JUL-1999; 99US-0144634.  
PR 21-JUL-1999; 99US-0144814.  
PR 21-JUL-1999; 99US-0145086.  
PR 21-JUL-1999; 99US-0145088.  
PR 22-JUL-1999; 99US-0145085.  
PR 22-JUL-1999; 99US-0145087.  
PR 22-JUL-1999; 99US-0145089.  
PR 22-JUL-1999; 99US-0145192.  
PR 23-JUL-1999; 99US-0145145.  
PR 23-JUL-1999; 99US-0145218.  
PR 23-JUL-1999; 99US-0145224.  
PR 26-JUL-1999; 99US-0145276.  
PR 27-JUL-1999; 99US-0145913.  
PR 27-JUL-1999; 99US-0145918.  
PR 27-JUL-1999; 99US-0145919.  
PR 28-JUL-1999; 99US-0145951.  
PR 02-AUG-1999; 99US-0146386.  
PR 02-AUG-1999; 99US-0146388.  
PR 02-AUG-1999; 99US-0146389.  
PR 03-AUG-1999; 99US-0147038.  
PR 04-AUG-1999; 99US-0147204.  
PR 04-AUG-1999; 99US-0147302.  
PR 05-AUG-1999; 99US-0147192.  
PR 05-AUG-1999; 99US-0147260.  
PR 06-AUG-1999; 99US-0147303.  
PR 06-AUG-1999; 99US-0147416.  
PR 09-AUG-1999; 99US-0147493.  
PR 09-AUG-1999; 99US-0147935.  
PR 10-AUG-1999; 99US-0148171.  
PR 11-AUG-1999; 99US-0148319.

PR 12-AUG-1999; 99US-0148341.  
PR 13-AUG-1999; 99US-0148565.  
PR 13-AUG-1999; 99US-0148684.  
PR 16-AUG-1999; 99US-0149368.  
PR 17-AUG-1999; 99US-0149175.  
PR 18-AUG-1999; 99US-0149426.  
PR 20-AUG-1999; 99US-0149722.  
PR 20-AUG-1999; 99US-0149723.  
PR 20-AUG-1999; 99US-0149929.  
PR 20-AUG-1999; 99US-0149902.  
PR 23-AUG-1999; 99US-0149930.  
PR 23-AUG-1999; 99US-0150566.  
PR 25-AUG-1999; 99US-0150884.  
PR 26-AUG-1999; 99US-0151065.  
PR 27-AUG-1999; 99US-0151066.  
PR 27-AUG-1999; 99US-0151080.  
PR 30-AUG-1999; 99US-0151303.  
PR 31-AUG-1999; 99US-0151438.  
PR 01-SEP-1999; 99US-0151930.  
PR 07-SEP-1999; 99US-0152363.  
PR 10-SEP-1999; 99US-0153070.  
PR 13-SEP-1999; 99US-0153758.  
PR 15-SEP-1999; 99US-0154018.  
PR 16-SEP-1999; 99US-0154039.  
PR 20-SEP-1999; 99US-0154779.  
PR 22-SEP-1999; 99US-0155139.  
PR 23-SEP-1999; 99US-0155486.  
PR 24-SEP-1999; 99US-0155659.  
PR 28-SEP-1999; 99US-0156458.  
PR 29-SEP-1999; 99US-0156596.  
PR 04-OCT-1999; 99US-0157117.  
PR 05-OCT-1999; 99US-0157753.  
PR 06-OCT-1999; 99US-0157865.  
PR 07-OCT-1999; 99US-0158029.  
PR 08-OCT-1999; 99US-0158232.  
PR 12-OCT-1999; 99US-0158369.  
PR 13-OCT-1999; 99US-0159293.  
PR 13-OCT-1999; 99US-0159295.  
PR 14-OCT-1999; 99US-0159329.  
PR 14-OCT-1999; 99US-0159330.  
PR 14-OCT-1999; 99US-0159331.  
PR 14-OCT-1999; 99US-0159637.  
PR 14-OCT-1999; 99US-0159638.  
PR 18-OCT-1999; 99US-0159584.  
PR 21-OCT-1999; 99US-0160741.  
PR 21-OCT-1999; 99US-0160767.  
PR 21-OCT-1999; 99US-0160768.  
PR 21-OCT-1999; 99US-0160770.  
PR 21-OCT-1999; 99US-0160814.  
PR 21-OCT-1999; 99US-0160815.  
PR 22-OCT-1999; 99US-0160980.  
PR 22-OCT-1999; 99US-0160981.  
PR 22-OCT-1999; 99US-0160989.  
PR 25-OCT-1999; 99US-0161404.  
PR 25-OCT-1999; 99US-0161405.  
PR 25-OCT-1999; 99US-0161406.  
PR 26-OCT-1999; 99US-0161359.  
PR 26-OCT-1999; 99US-0161360.  
PR 26-OCT-1999; 99US-0161361.  
PR 28-OCT-1999; 99US-0161920.  
PR 28-OCT-1999; 99US-0161992.  
PR 28-OCT-1999; 99US-0161993.  
PR 29-OCT-1999; 99US-0162142.

Query Match 2.5%; Score 10; DB 21; Length 240;  
Best Local Similarity 100.0%; Pred. No. 0.53;  
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 48 pppppppppn 57  
Db 84 pppppppppn 93

RESULT 12  
AAG39267  
ID AAG39267 standard; Protein; 263 AA.  
XX  
AC AAG39267;  
XX  
DT 18-OCT-2000 (first entry)  
XX  
DE Arabidopsis thaliana protein fragment SEQ ID NO: 48561.  
XX  
KW Protein identification; signal transduction pathway; metabolic pathway;  
KW hybridisation assay; genetic mapping; gene expression control; promoter;  
KW termination sequence.  
XX  
OS Arabidopsis thaliana.  
XX  
PN EP1033405-A2.  
XX  
PD 06-SEP-2000.  
XX  
PF 25-FEB-2000; 2000EP-0301439.  
XX  
PR 25-FEB-1999; 99US-0121825.  
PR 05-MAR-1999; 99US-0123180.  
PR 09-MAR-1999; 99US-0123548.  
PR 23-MAR-1999; 99US-0125788.  
PR 25-MAR-1999; 99US-0126284.  
PR 29-MAR-1999; 99US-0126785.  
PR 01-APR-1999; 99US-0127462.  
PR 06-APR-1999; 99US-0128234.  
PR 08-APR-1999; 99US-0128714.  
PR 16-APR-1999; 99US-0129845.  
PR 19-APR-1999; 99US-0130077.  
PR 21-APR-1999; 99US-0130449.  
PR 23-APR-1999; 99US-0130510.  
PR 23-APR-1999; 99US-0130891.  
PR 28-APR-1999; 99US-0131449.  
PR 30-APR-1999; 99US-0132048.  
PR 30-APR-1999; 99US-0132407.  
PR 04-MAY-1999; 99US-0132484.  
PR 05-MAY-1999; 99US-0132485.  
PR 06-MAY-1999; 99US-0132486.  
PR 06-MAY-1999; 99US-0132487.  
PR 07-MAY-1999; 99US-0132863.  
PR 11-MAY-1999; 99US-0134256.  
PR 14-MAY-1999; 99US-0134218.  
PR 14-MAY-1999; 99US-0134219.  
PR 14-MAY-1999; 99US-0134221.  
PR 14-MAY-1999; 99US-0134370.  
PR 18-MAY-1999; 99US-0134768.  
PR 19-MAY-1999; 99US-0134941.  
PR 20-MAY-1999; 99US-0135124.  
PR 21-MAY-1999; 99US-0135353.  
PR 24-MAY-1999; 99US-0135629.  
PR 25-MAY-1999; 99US-0136021.  
PR 27-MAY-1999; 99US-0136392.  
PR 28-MAY-1999; 99US-0136782.  
PR 01-JUN-1999; 99US-0137222.  
PR 03-JUN-1999; 99US-0137528.  
PR 04-JUN-1999; 99US-0137502.  
PR 07-JUN-1999; 99US-0137724.  
PR 08-JUN-1999; 99US-0138094.  
PR 10-JUN-1999; 99US-0138540.  
PR 10-JUN-1999; 99US-0138847.  
PR 14-JUN-1999; 99US-0139119.  
PR 16-JUN-1999; 99US-0139452.  
PR 16-JUN-1999; 99US-0139453.  
PR 17-JUN-1999; 99US-0139492.  
PR 18-JUN-1999; 99US-0139454.  
PR 18-JUN-1999; 99US-0139455.  
PR 18-JUN-1999; 99US-0139456.  
PR 18-JUN-1999; 99US-0139457.  
PR 18-JUN-1999; 99US-0139458.  
PR 18-JUN-1999; 99US-0139459.  
PR 18-JUN-1999; 99US-0139460.  
PR 18-JUN-1999; 99US-0139461.  
PR 18-JUN-1999; 99US-0139462.  
PR 18-JUN-1999; 99US-0139463.  
PR 18-JUN-1999; 99US-0139750.  
PR 18-JUN-1999; 99US-0139763.  
PR 21-JUN-1999; 99US-0139817.  
PR 22-JUN-1999; 99US-0139899.  
PR 23-JUN-1999; 99US-0140353.  
PR 23-JUN-1999; 99US-0140354.  
PR 24-JUN-1999; 99US-0140695.  
PR 28-JUN-1999; 99US-0140823.  
PR 29-JUN-1999; 99US-0140991.  
PR 30-JUN-1999; 99US-0141287.  
PR 01-JUL-1999; 99US-0141842.  
PR 01-JUL-1999; 99US-0142154.  
PR 02-JUL-1999; 99US-0142055.  
PR 06-JUL-1999; 99US-0142390.  
PR 08-JUL-1999; 99US-0142803.  
PR 09-JUL-1999; 99US-0142920.  
PR 12-JUL-1999; 99US-0142977.  
PR 13-JUL-1999; 99US-0143542.  
PR 14-JUL-1999; 99US-0143624.  
PR 15-JUL-1999; 99US-0144005.  
PR 16-JUL-1999; 99US-0144085.  
PR 16-JUL-1999; 99US-0144086.  
PR 19-JUL-1999; 99US-0144325.  
PR 19-JUL-1999; 99US-0144331.  
PR 19-JUL-1999; 99US-0144332.  
PR 19-JUL-1999; 99US-0144333.  
PR 19-JUL-1999; 99US-0144334.  
PR 19-JUL-1999; 99US-0144335.  
PR 20-JUL-1999; 99US-0144352.  
PR 20-JUL-1999; 99US-0144632.  
PR 20-JUL-1999; 99US-0144884.  
PR 21-JUL-1999; 99US-0144814.  
PR 21-JUL-1999; 99US-0145086.  
PR 21-JUL-1999; 99US-0145088.  
PR 22-JUL-1999; 99US-0145085.  
PR 22-JUL-1999; 99US-0145087.  
PR 22-JUL-1999; 99US-0145089.  
PR 22-JUL-1999; 99US-0145192.  
PR 23-JUL-1999; 99US-0145145.  
PR 23-JUL-1999; 99US-0145218.  
PR 23-JUL-1999; 99US-0145224.  
PR 26-JUL-1999; 99US-0145276.  
PR 27-JUL-1999; 99US-0145913.  
PR 27-JUL-1999; 99US-0145918.  
PR 27-JUL-1999; 99US-0145919.  
PR 28-JUL-1999; 99US-0145951.  
PR 02-AUG-1999; 99US-0146386.  
PR 02-AUG-1999; 99US-0146388.  
PR 02-AUG-1999; 99US-0146389.  
PR 03-AUG-1999; 99US-0147038.  
PR 04-AUG-1999; 99US-0147204.  
PR 04-AUG-1999; 99US-0147302.  
PR 05-AUG-1999; 99US-0147192.  
PR 05-AUG-1999; 99US-0147260.  
PR 06-AUG-1999; 99US-0147303.  
PR 06-AUG-1999; 99US-0147416.  
PR 09-AUG-1999; 99US-0147493.  
PR 09-AUG-1999; 99US-0147494.  
PR 09-AUG-1999; 99US-0147935.  
PR 10-AUG-1999; 99US-0148171.  
PR 11-AUG-1999; 99US-0148319.  
PR 12-AUG-1999; 99US-0148341.  
PR 13-AUG-1999; 99US-0148565.  
PR 13-AUG-1999; 99US-0148684.  
PR 16-AUG-1999; 99US-0149368.  
PR 17-AUG-1999; 99US-0149175.  
PR 18-AUG-1999; 99US-0149426.  
PR 20-AUG-1999; 99US-0149722.

PR 20-AUG-1999; 99US-0149723.  
PR 20-AUG-1999; 99US-0149929.  
PR 23-AUG-1999; 99US-0149902.  
PR 23-AUG-1999; 99US-0149930.  
PR 25-AUG-1999; 99US-0150566.  
PR 26-AUG-1999; 99US-0150884.  
PR 27-AUG-1999; 99US-0151065.  
PR 27-AUG-1999; 99US-0151066.  
PR 27-AUG-1999; 99US-0151080.  
PR 30-AUG-1999; 99US-0151303.  
PR 31-AUG-1999; 99US-0151438.  
PR 01-SEP-1999; 99US-0151930.  
PR 07-SEP-1999; 99US-0152363.  
PR 10-SEP-1999; 99US-0153070.  
PR 13-SEP-1999; 99US-0153758.  
PR 15-SEP-1999; 99US-0154018.  
PR 16-SEP-1999; 99US-0154039.  
PR 20-SEP-1999; 99US-0154779.  
PR 22-SEP-1999; 99US-0155139.  
PR 23-SEP-1999; 99US-0155486.  
PR 24-SEP-1999; 99US-0155659.  
PR 28-SEP-1999; 99US-0156458.  
PR 29-SEP-1999; 99US-0156596.  
PR 04-OCT-1999; 99US-0157117.  
PR 03-OCT-1999; 99US-0157753.  
PR 06-OCT-1999; 99US-0157865.  
PR 07-OCT-1999; 99US-0158029.  
PR 08-OCT-1999; 99US-0158232.  
PR 12-OCT-1999; 99US-0158369.  
PR 13-OCT-1999; 99US-0159293.  
PR 13-OCT-1999; 99US-0159294.  
PR 13-OCT-1999; 99US-0159295.  
PR 14-OCT-1999; 99US-0159329.  
PR 14-OCT-1999; 99US-0159330.  
PR 14-OCT-1999; 99US-0159331.  
PR 14-OCT-1999; 99US-0159637.  
PR 14-OCT-1999; 99US-0159638.  
PR 18-OCT-1999; 99US-0159584.  
PR 21-OCT-1999; 99US-0160741.  
PR 21-OCT-1999; 99US-0160767.  
PR 21-OCT-1999; 99US-0160768.  
PR 21-OCT-1999; 99US-0160770.  
PR 21-OCT-1999; 99US-0160814.  
PR 21-OCT-1999; 99US-0160815.  
PR 22-OCT-1999; 99US-0160980.  
PR 22-OCT-1999; 99US-0160981.  
PR 22-OCT-1999; 99US-0160989.  
PR 25-OCT-1999; 99US-0161404.  
PR 25-OCT-1999; 99US-0161405.  
PR 25-OCT-1999; 99US-0161406.  
PR 26-OCT-1999; 99US-0161359.  
PR 26-OCT-1999; 99US-0161360.  
PR 26-OCT-1999; 99US-0161361.  
PR 28-OCT-1999; 99US-0161820.  
PR 28-OCT-1999; 99US-0161992.  
PR 28-OCT-1999; 99US-0161993.  
PR 29-OCT-1999; 99US-0162142.

Query Match 2.5%; Score 10; DB 21; Length 263;  
Best Local Similarity 100.0%; Pred. No. 0.57; 0; Indels 0; Gaps 0;  
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 48 pppppppppn 57  
|||||  
Db 107 pppppppppn 116

RESULT 13  
ABG25300  
ID ABG25300 standard; Protein; 697 AA.  
XX  
AC ABG25300;  
XX

DT 18-FEB-2002 (first entry)  
XX  
DE Novel human diagnostic protein #25291.  
XX  
XX Human; chromosome mapping; gene mapping; gene therapy; forensic;  
KW food supplement; medical imaging; diagnostic; genetic disorder.  
XX  
XX Homo sapiens.  
XX  
XX WO200175067-A2.  
XX  
XX 11-OCT-2001.  
XX  
XX 30-MAR-2001; 2001WO-US08631.  
XX  
XX 31-MAR-2000; 2000US-0540217.  
PR 23-AUG-2000; 2000US-0649167.  
XX  
XX (HYSE-) HYSEQ INC.  
XX  
XX Drmanac RT, Liu C, Tang YT;  
XX  
XX WPI; 2001-639362/73.  
DR N-PSDB; AAS89487.  
DR  
XX  
XX New isolated polynucleotide and encoded polypeptides, useful in  
PT diagnostics, forensics, gene mapping, identification of mutations  
PT responsible for genetic disorders or other traits and to assess  
PT biodiversity  
XX  
XX Claim 20; SEQ ID No 55659; 103pp; English.  
XX  
XX The invention relates to isolated polynucleotide (I) and  
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,  
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome  
CC and gene mapping, and in recombinant production of (II). The  
CC polynucleotides are also used in diagnostics as expressed sequence tags  
CC for identifying expressed genes. (I) is useful in gene therapy techniques  
CC to restore normal activity of (II) or to treat disease states involving  
CC (II). (II) is useful for generating antibodies against it, detecting or  
CC quantitating a polypeptide in tissue, as molecular weight markers and as  
CC a food supplement. (II) and its binding partners are useful in medical  
CC imaging of sites expressing (II). (I) and (II) are useful for treating  
CC disorders involving aberrant protein expression or biological activity.  
CC The polypeptide and polynucleotide sequences have applications in  
CC diagnostics, forensics, gene mapping, identification of mutations  
CC responsible for genetic disorders or other traits to assess biodiversity  
CC and to produce other types of data and products dependent on DNA and  
CC amino acid sequences. ABG00010-ABG30377 represent novel human  
CC diagnostic amino acid sequences of the invention.  
CC Note: The sequence data for this patent did not appear in the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pct\_sequences.

XX Sequence 697 AA;

Query Match 2.5%; Score 10; DB 22; Length 697;  
Best Local Similarity 100.0%; Pred. No. 1.3;  
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 140 AAAAAAPPPP 149  
|||||  
Db 520 aaaaaapppp 529

RESULT 14  
AAP20304  
ID AAP20304 standard; peptide; 9 AA.  
XX  
AC AAP20304;  
XX  
DT 30-NOV-1992 (first entry)

xx DE Peptide for use in an anti-wrinkle cosmetic.  
xx KW Cream; non-irritating.  
xx XX Synthetic.  
xx PN JP57002214-A.  
xx PD 07-JAN-1982.  
xx PF 06-JUN-1980; 80JP-0076267.  
xx PR 05-JUN-1980; 80JP-0076268.  
xx PA (MTU ) MITSUBISHI CHEM IND KK.  
xx PA (KANE-) KANEBO KESHOIN KK.  
xx DR WPI; 1982-12597E/07 (12597E).  
xx PT Peptide-contg. cosmetic - preventing formation of fine wrinkles  
xx PS Claim 1; Page 4; 6pp; Japanese.  
xx CC The peptide is used in a cosmetic to prevent the formation of fine  
CC wrinkles without irritating the skin. The peptide has at its N-  
CC terminus an organic gp. which is capable of reacting with hydrogen  
CC or amino gps. e.g. a benzoyl, acetyl, formyl, tosyl gp. etc. The  
CC peptide may be prepd. by standard peptide synthesis techniques.  
xx SQ Sequence 9 AA;  
  
Query Match 2.2%; Score 9; DB 3; Length 9;  
Best Local Similarity 100.0%; Pred. No. 6.4e+05;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
Qy 48 pppppppppp 56  
Db 1 pppppppppp 9  
  
RESULT 15  
AAU09969  
ID AAU09969 standard; peptide; 10 AA.  
XX AC AAU09969;  
XX DT 16-JAN-2002 (first entry)  
XX DE Proline linker sequence.  
XX KW Cancer; caspase; apoptosis; Alzheimer's disease;  
KW Down's syndrome; prion protein; Bovine Spongiform Encephalopathy;  
KW BSE; acquired immune deficiency syndrome; AIDS; proline linker;  
KW Creutzfeld-Jacob disease; CJD.  
XX OS Synthetic.  
XX WO200175453-A2.  
XX PD 11-OCT-2001.  
XX PF 04-APR-2001; 2001WO-GB01540.  
XX PR 04-APR-2000; 2000GB-0008254.  
XX PR 04-APR-2000; 2000GB-0008256.  
XX PA (MEDI-) MEDICAL RES COUNCIL.  
XX PI Tse E, Rabbitts T;  
XX DR WPI; 2001-626485/72.

xx XX Inducing cell to generate detectable signal and detecting entity within  
PT cell, comprises allowing stable interaction of first reporter and  
PT second reporter by binding with the entity and generating signal -  
xx PS Disclosure; Page 45; 76pp; English.  
xx XX  
CC This sequence represents a proline linker sequence used for cloning  
CC F-box proteins from mammalian cells or tissues in the method of  
CC the invention. The invention comprises a method of inducing a cell to  
CC generate a detectable signal and detecting an entity within a cell. The  
CC method involves providing a cell containing an entity with a first and  
CC second reporter where a stable interaction between the reporters  
CC leads to generation of a detectable signal which can be monitored. Also  
CC included is a method for killing a cell in which a cell containing a  
CC given entity is given a first reporter containing a first immunoglobulin  
CC (Igl) and a second reporter containing a second immunoglobulin (Ig2).  
CC The Igl molecule is linked to a first caspase molecule and a second  
CC caspase molecule is linked to Ig2, where the two caspases are capable of  
CC stably interacting to generate caspase activity and induce apoptosis.  
CC This method can also be used for destruction of a polypeptide in a cell,  
CC by providing a cell containing a polypeptide with two reporters where a  
CC stable interaction between these reporters leads to generation of  
CC protease activity, and proteolysis of the polypeptide. The function of  
CC unknown genes can be determined with this method by using this technique  
CC to destroy the polypeptide encoded by the gene and observing a  
CC phenotype. These methods are useful for detecting an entity within a cell  
CC (e.g. senile plaques in Alzheimer's disease, neurofibrillary tangles in  
CC Down's syndrome or infectious prion proteins associated with Bovine  
CC Spongiform Encephalopathy (BSE), Creutzfeld-Jacob disease (CJD) or  
CC proteins associated with acquired immune deficiency syndrome (AIDS)),  
CC killing a cell, destruction of a polypeptide in a cell, functional  
CC genomic studies and in the treatment or diagnosis of a cancer.  
xx SQ Sequence 10 AA;  
  
Query Match 2.2%; Score 9; DB 22; Length 10;  
Best Local Similarity 100.0%; Pred. No. 0.28;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
Qy 48 pppppppppp 56  
Db 1 pppppppppp 9  
  
Search completed: August 13, 2002, 15:13:36  
Job time: 142 sec



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 13, 2002, 15:12:04 ; Search time 20.85 Seconds  
(without alignments)  
1848.049 Million cell updates/sec

Title: US-09-631-863A-2  
Perfect score: 401  
Sequence: 1 MNKRVLTKKRRSGRGQD.....DNCPAKASKSPAGNSAPL 401

Scoring table: OLIGO  
Gapop 60.0 , Gapext 60.0

Searched: 283138 seqs, 96089334 residues

Word size : 0

Total number of hits satisfying chosen parameters: 283138

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database :  
1: pir1:\*  
2: pir2:\*  
3: pir3:\*  
4: pir4:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB ID	Description
1	15	3.7	232	2 JE0163	myelin expression
2	10	2.5	250	2 T16342	hypothetical prote
3	10	2.5	494	2 B96534	hypothetical prote
4	10	2.5	1018	2 T43168	hypothetical prote
5	10	2.5	1259	2 T16038	hypothetical prote
6	9	2.2	90	2 T18068	hypothetical prote
7	9	2.2	135	2 E96750	hypothetical prote
8	9	2.2	141	2 A34043	hypothetical proli
9	9	2.2	142	2 B41132	collagen-related p
10	9	2.2	148	2 S39206	proline-rich prote
11	9	2.2	149	2 A41132	collagen-related p
12	9	2.2	161	2 H85120	probable proline-r
13	9	2.2	165	2 C71717	hypothetical prote
14	9	2.2	172	2 D41132	collagen-related p
15	9	2.2	172	2 T27505	hypothetical prote
16	9	2.2	182	2 T30760	hypothetical prote
17	9	2.2	186	2 C41132	collagen-related p
18	9	2.2	195	2 T07735	nodulin-20a - soyb
19	9	2.2	219	2 T10563	proline-rich prote
20	9	2.2	220	2 S52005	nodulin-30 (Np30)
21	9	2.2	221	2 H82857	tonB protein XF000
22	9	2.2	229	1 W4WL8	E4 protein - human
23	9	2.2	236	2 T02577	probable AP2 domai
24	9	2.2	246	2 T46446	hypothetical prote
25	9	2.2	250	1 A31757	homeotic protein H
26	9	2.2	250	2 T08908	hypothetical prote
27	9	2.2	251	1 B60492	homeotic protein H
28	9	2.2	255	2 C36222	phosphoprotein pho
29	9	2.2	257	2 T00406	hypothetical prote

ALIGNMENTS

RESULT 1

JE0163  
myelin expression factor-3 - mouse  
N:Alternate names: MYEF-3  
C:Species: Mus musculus (house mouse)  
C:Date: 03-Jul-1998 #sequence\_revision 10-Jul-1998 #text\_change 07-May-1999  
C:Accession: JE0163  
R:Stepiowski, A.; Krynska, B.; Tretiakova, A.; Haas, S.; Khailli, K.; Amini, S.  
Biochem. Biophys. Res. Commun. 243, 295-301, 1998  
A:Title: MYEF-3, a developmentally controlled brain-derived nuclear protein which spe  
A:Reference number: JE0163; MUID:98139908  
A:Accession: JE0163  
A:Molecule type: mRNA  
A:Residues: 1-232 <STE>  
A:Experimental source: brain  
C:Keywords: phosphoprotein  
F:40-60/Domain: transmembrane #status predicted <TM>  
F:31,175/Binding site: phosphate (Ser) (covalent) (by protein kinase C) #status predi  
F:31,142,185,201/Binding site: phosphate (Ser) (covalent) (by casein kinase II) #stat

Query Match 3.7%; Score 15; DB 2; Length 232;  
Best Local Similarity 100.0%; Pred. No. 7.1e-06;  
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 259 DYSNAFQMIQDLQDW 273

Db 106 DYSNAFQMIQDLQDW 120

RESULT 2

T16342  
hypothetical protein F42C5.7 - Caenorhabditis elegans  
C:Species: Caenorhabditis elegans  
C:Date: 20-Sep-1999 #sequence\_revision 20-Sep-1999 #text\_change 20-Sep-1999  
C:Accession: T16342  
R:Du, Z.  
submitted to the EMBL Data Library, November 1995  
A:Description: The sequence of C. elegans cosmid F42C5.  
A:Reference number: Z18497  
A:Accession: T16342  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-250 <DUZ>  
A:Cross-references: EMBL:U40799; NID:g1065935; PID:g1065941; PIDN:AAA81484.1; CESP:F4  
C:Genetics:  
A:Gene: CESP:F42C5.7  
A:Introns: 15/2; 40/3; 167/3; 191/3

Query Match 2.5%; Score 10; DB 2; Length 250;  
Best Local Similarity 100.0%; Pred. No. 0.29;

```

Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 47 C P P P P P P P P P 56
      |||||
Db 102 C P P P P P P P P P 111

RESULT 3
B96534
hypothetical protein F14J22.4 [imported] - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 31-Mar-2001
C:Accession: B96534
R:Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso,
Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Creasy, T.H.; Dewar, K.;
ansen, N.F.; Hughes, B.; Huizar, L.
Nature 408, 816-820, 2000
A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.
C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Maiti, R.; Marziali,
Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.
A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon,
ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.
A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.
A:Reference number: A86141; MUID:21016719
A:Accession: B96534
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-494 <STO>
A:Cross-references: GB:AE005173; NID:g10120434; PIDN:AGI3059.1; GSPDB:GN00141
C:Genetics:
A:Gene: F14J22.4
A:Map position: 1

Query Match 2.5%; Score 10; DB 2; Length 494;
Best Local Similarity 100.0%; Pred. No. 0.52;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 46 D C P P P P P P P P 55
      |||||
Db 60 D C P P P P P P P P 69

RESULT 4
T43168
hypothetical protein cyk-1 - Caenorhabditis elegans (fragment)
C:Species: Caenorhabditis elegans
C:Date: 09-Jun-2000 #sequence_revision 09-Jun-2000 #text_change 09-Jun-2000
C:Accession: T43168
R:Swan, K.A.; Severson, A.F.; Carter, J.C.; Martin, P.R.; Schnabel, H.; Schnabel, R.; Bo
J. Cell Sci. 111, 2017-2027, 1998
A:Title: cyk-1: a C. elegans FH gene required for a late step in embryonic cytokinesis.
A:Reference number: 222311; MUID:98311662
A:Accession: T43168
A:Status: preliminary; translated from GB/EMBL/DDBJ
A:Molecule type: mRNA
A:Residues: 1-1018 <SWA>
A:Cross-references: EMBL:AF062008; PIDN:AAC17501.1
C:Genetics:
A:Gene: cyk-1
A:Map position: 3

Query Match 2.5%; Score 10; DB 2; Length 1018;
Best Local Similarity 100.0%; Pred. No. 0.97;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 47 C P P P P P P P P P 56
      |||||
Db 351 C P P P P P P P P P 360

RESULT 5

```

```

T16038
hypothetical protein F11H8.4 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 20-Sep-1999
C:Accession: T16038
R:Menezes, S.
submitted to the EMBL Data Library, November 1995
A:Description: The sequence of C. elegans cosmid F11H8.
A:Reference number: Z18450
A:Accession: T16038
A:Status: preliminary; translated from GB/EMBL/DDBJ
A:Molecule type: DNA
A:Residues: 1-1259 <MEN>
A:Cross-references: EMBL:U40187; NID:g1055194; PID:g1055198; PIDN:AAA81161.1; CESP:F1
C:Genetics:
A:Gene: CESP:F11H8.4
A:Introns: 160/3; 192/1; 212/3; 291/3; 350/3; 417/2; 511/3; 632/1; 677/3; 895/1; 1073

Query Match 2.5%; Score 10; DB 2; Length 1259;
Best Local Similarity 100.0%; Pred. No. 1.2;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 47 C P P P P P P P P P 56
      |||||
Db 768 C P P P P P P P P P 777

RESULT 6
T18068
hypothetical protein a566L - Chlorella virus PBCV-1
C:Species: Chlorella virus PBCV-1
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
C:Accession: T18068
R:Graves, M.V.; Van Etten, J.L.
submitted to the EMBL data Library, May 1999
A:Reference number: Z18806
A:Accession: T18068
A:Status: preliminary; translated from GB/EMBL/DDBJ
A:Molecule type: DNA
A:Residues: 1-90 <GRA>
A:Cross-references: EMBL:U42580; NID:g4028896; PIDN:AAC97003.1
A:Experimental source: specific host Chlorella strain NC64A
C:Genetics:
A:Note: a566L

Query Match 2.2%; Score 9; DB 2; Length 90;
Best Local Similarity 100.0%; Pred. No. 0.98;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 48 P P P P P P P P P P 56
      |||||
Db 39 P P P P P P P P P P 47

RESULT 7
E96750
hypothetical protein F28P22.21 [imported] - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 31-Mar-2001
C:Accession: E96750
R:Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alon
Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Creasy, T.H.; Dewar,
ansen, N.F.; Hughes, B.; Huizar, L.
Nature 408, 816-820, 2000
A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim,
C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Maiti, R.; Marzia
Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.
A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallo
ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.
A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.
A:Reference number: A86141; MUID:21016719

```



A;Accession: E96750  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-135 <STO>  
A;Cross-references: GB:AE0051173; NID:96648168; PIDN:AAF21168.1; GSPDB:GN00141  
C;Genetics:  
A;Map position: 1

Query Match 2.2%; Score 9; DB 2; Length 135;  
Best Local Similarity 100.0%; Pred. No. 1.4;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 48 PPPPPPPPP 56  
|||||

Db 38 PPPPPPPPP 46  
|||||

RESULT 8  
A34043  
hypothetical proline-rich protein 1 - polychaete (Owenia fusiformis) (fragment)  
C;Species: Owenia fusiformis  
C;Date: 07-Jun-1990 #sequence\_revision 07-Jun-1990 #text\_change 01-Dec-2000  
C;Accession: A34043; B34043  
R;Bakalara, N.; Collet, J.; Planells, R.; Thouveny, Y.; Fontes, M.  
Biochem. Biophys. Res. Commun. 166, 66-73, 1990  
A;Title: Presence in invertebrate genomes of sequences characterized by the repetition  
A;Reference number: A90159; MUID:90147742  
A;Accession: A34043  
A;Molecule type: DNA  
A;Residues: 1-141 <BAK>  
A;Cross-references: GB:M32217  
A;Accession: B34043  
A;Molecule type: DNA  
A;Residues: 59-136 <BA2>  
A;Cross-references: GB:M32217

Query Match 2.2%; Score 9; DB 2; Length 141;  
Best Local Similarity 100.0%; Pred. No. 1.4;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 48 PPPPPPPPP 56  
|||||

Db 9 PPPPPPPPP 17  
|||||

RESULT 9  
B41132  
collagen-related protein 2 - Hydra magnipapillata (fragment)  
C;Species: Hydra magnipapillata  
C;Date: 05-Jun-1992 #sequence\_revision 05-Jun-1992 #text\_change 01-Dec-2000  
C;Accession: B41132; S21930  
R;Kurz, E.M.; Holstein, T.W.; Petri, B.M.; Engel, J.; David, C.N.  
J. Cell Biol. 115, 1159-1169, 1991  
A;Title: Mini-collagens in hydra nematocytes.  
A;Reference number: A41132; MUID:92064646  
A;Accession: B41132  
A;Status: preliminary  
A;Molecule type: mRNA  
A;Residues: 1-142 <KUR>  
A;Cross-references: EMBL:X61046; NID:99448; PIDN:CAA43380.1; PID:99449  
A;Note: submitted to the EMBL Data Library, July 1991  
C;Superfamily: unassigned collagens

Query Match 2.2%; Score 9; DB 2; Length 142;  
Best Local Similarity 100.0%; Pred. No. 1.5;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 48 PPPPPPPPP 56  
|||||

Db 46 PPPPPPPPP 54  
RESULT 10  
S39206  
proline-rich protein V-beta 1 precursor - rat  
C;Species: Rattus norvegicus (Norway rat)  
C;Date: 06-Jan-1995 #sequence\_revision 06-Jan-1995 #text\_change 19-May-2000  
C;Accession: A53118; S39206  
R;Court, Y.; Rosinski-Chupin, I.; Rougeon, F.  
J. Biol. Chem. 269, 520-527, 1994  
A;Title: A new proline-rich protein precursor expressed in the salivary glands of the  
A;Reference number: A53118; MUID:94103265  
A;Accession: A53118  
A;Status: translated from GB/EMBL/DBJ  
A;Molecule type: mRNA  
A;Residues: 1-148 <COU>  
A;Cross-references: EMBL:X74229; NID:9433616; PIDN:CAA52300.1; PID:9433617  
A;Note: submitted to the EMBL Data Library, July 1993  
C;Genetics:  
A;Gene: VCS-beta1  
A;Introns: 18/3  
C;Superfamily: proline-rich peptide P-B  
C;Keywords: glycoprotein  
F;1-18/Domain: signal sequence #status predicted <SIG>  
F;19-148/Product: proline-rich protein V-beta 1 #status predicted <WAT>  
F;133,143/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 2.2%; Score 9; DB 2; Length 148;  
Best Local Similarity 100.0%; Pred. No. 1.5;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 48 PPPPPPPPP 56  
|||||

Db 35 PPPPPPPPP 43  
|||||

RESULT 11  
A41132  
collagen-related protein 1 precursor - Hydra magnipapillata  
C;Species: Hydra magnipapillata  
C;Date: 05-Jun-1992 #sequence\_revision 05-Jun-1992 #text\_change 01-Dec-2000  
C;Accession: A41132; S21929  
R;Kurz, E.M.; Holstein, T.W.; Petri, B.M.; Engel, J.; David, C.N.  
J. Cell Biol. 115, 1159-1169, 1991  
A;Title: Mini-collagens in hydra nematocytes.  
A;Reference number: A41132; MUID:92064646  
A;Accession: A41132  
A;Status: preliminary  
A;Molecule type: mRNA  
A;Residues: 1-149 <KUR>  
A;Cross-references: EMBL:X61045; NID:99446; PIDN:CAA43379.1; PID:99447  
A;Note: submitted to the EMBL Data Library, July 1991  
C;Superfamily: unassigned collagens

Query Match 2.2%; Score 9; DB 2; Length 149;  
Best Local Similarity 100.0%; Pred. No. 1.5;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 48 PPPPPPPPP 56  
|||||

Db 52 PPPPPPPPP 60  
|||||

RESULT 12  
H85120  
probable proline-rich protein [imported] - Arabidopsis thaliana  
C;Species: Arabidopsis thaliana (mouse-ear cress)  
C;Date: 16-Feb-2001 #sequence\_revision 16-Feb-2001 #text\_change 16-Feb-2001  
C;Accession: H85120  
R;anonymous, The European Union Arabidopsis Genome Sequencing Consortium, The Cold Sp

Nature 402, 769-777, 1999  
 A:Title: Sequence and analysis of chromosome 4 of the plant Arabidopsis thaliana.  
 A:Reference number: A85001; MUID:20083488

A:Accession: H85120  
 A>Status: Preliminary

A:Molecule type: DNA

A:Residues: 1-161 <STOS>

A:Cross-references: GB:NC\_001268; NID:g7267842; PIDN:CAB81244.1; GSPDB:GN00140

C:Genetics:

A:Gene: AT4g11430

A:Map position: 4

#### Query Match

Best Local Similarity 2.2%; Score 9; DB 2; Length 161;

Best Local Similarity 100.0%; Pred. No. 1.6;

Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 48 PPPPPPPPP 56

|||||

DB 120 PPPPPPPPP 128

#### RESULT 13

C71717  
 hypothetical protein RP084 - Rickettsia prowazekii

C:Species: Rickettsia prowazekii

C:Date: 21-Nov-1998 #sequence\_revision 21-Nov-1998 #text\_change 03-Nov-2000

C:Accession: C71717

R:Andersson, S.G.E.; Zomorodipour, A.; Andersson, J.O.; Sicheritz-Ponten, T.; Alsmark, U.

Nature 396, 133-140, 1998

A:Title: The genome sequence of Rickettsia prowazekii and the origin of mitochondria.

A:Reference number: A71630; MUID:99039499

A:Accession: C71717

A>Status: Preliminary; nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-165 <AND>

A:Cross-references: GB:AJ235270; GB:AJ235269; NID:g3860572; PIDN:CAA14554.1; PID:el34239

A:Experimental source: strain Madrid E

C:Genetics:

A:Gene: RP084

#### Query Match

Best Local Similarity 2.2%; Score 9; DB 2; Length 165;

Best Local Similarity 100.0%; Pred. No. 1.7;

Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 48 PPPPPPPPP 56

|||||

DB 36 PPPPPPPPP 44

#### RESULT 14

D41132

collagen-related protein 4 - Hydra magnipapillata (fragment)

C:Species: Hydra magnipapillata

C:Date: 05-Jun-1992 #sequence\_revision 12-Jun-1992 #text\_change 01-Dec-2000

C:Accession: D41132; S21932

R:Kurz, E.M.; Holstein, T.W.; Petri, B.M.; Engel, J.; David, C.N.

J. Cell Biol. 115, 1159-1169, 1991

A:Title: Mini-collagens in hydra nematocytes.

A:Reference number: A41132; MUID:92064646

A:Accession: D41132

A>Status: Preliminary

A:Molecule type: mRNA

A:Residues: 1-172 <KUR>

A:Cross-references: EMBL:X61048; NID:g9452; PIDN:CAA43382.1; PID:g9453

A:Note: the authors translated the codon GGT for residue 142 as Pro, and TTA for residue

A:Note: submitted to the EMBL Data Library, July 1991

C:Superfamily: unassigned collagens

#### Query Match

Best Local Similarity 2.2%; Score 9; DB 2; Length 172;

Best Local Similarity 100.0%; Pred. No. 1.7;

Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 48 PPPPPPPPP 56

|||||

DB 51 PPPPPPPPP 59

#### RESULT 15

T27505

hypothetical protein ZC168.5 - Caenorhabditis elegans

C:Species: Caenorhabditis elegans

C:Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 15-Oct-1999

C:Accession: T27505

R:Berks, M.

submitted to the EMBL Data Library, March 1996

A:Reference number: Z20378

A:Accession: T27505

A>Status: Preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-172 <WIL>

A:Cross-references: EMBL:Z70312; PIDN:CAA94385.1; GSPDB:GN00022; CESP:ZC168.5

A:Experimental source: clone ZC168

C:Genetics:

A:Gene: CESP:ZC168.5

A:Map position: 4

A:Introns: 112/3

#### Query Match

Best Local Similarity 2.2%; Score 9; DB 2; Length 172;

Best Local Similarity 100.0%; Pred. No. 1.7;

Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 47 CPPPPPPPP 55

|||||

DB 38 CPPPPPPPP 46

Search completed: August 13, 2002, 15:14:26

Job time: 142 sec



**THIS PAGE BLANK (USPTO)**

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 13, 2002, 15:13:04 ; Search time 13.48 Seconds  
(without alignments)  
1151.820 Million cell updates/sec

Title: US-09-631-863A-2  
Perfect score: 401  
Sequence: 1 MRNRVLTKRRSRGRGQD.....DNCAPAKASKSPAGNSAPPL 401

Scoring table: OLIGO  
Gapop 60.0 , Gapext 60.0

Searched: 105224 seqs, 38719550 residues

Word size : 0

Total number of hits satisfying chosen parameters: 105224

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : SwissProt\_40.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	10	2.5	387	1 UL33_RCMVM	O12000 rat cytochrome
2	9	2.2	141	1 YPRO_OWEFU	P21260 owenia fusi
3	9	2.2	165	1 Y084_RICPR	Q9ze63 rickettsia
4	9	2.2	223	1 TONB_XANCP	O34261 xanthomonas
5	9	2.2	229	1 VE4_HPV08	P06425 human papil
6	9	2.2	250	1 HX84_MOUSE	P10284 mus musculu
7	9	2.2	251	1 HX84_HUMAN	P17483 homo sapien
8	9	2.2	272	1 TRPC_MYCTU	O06129 mycobacteri
9	9	2.2	281	1 FASL_HUMAN	P48023 homo sapien
10	9	2.2	287	1 SMN_BOVIN	O18870 bos taurus
11	9	2.2	289	1 SMN_RAT	O35876 rattus norv
12	9	2.2	290	1 HX88_HUMAN	P13378 homo sapien
13	9	2.2	294	1 SMN_HUMAN	Q16637 homo sapien
14	9	2.2	297	1 HX1M_CHICK	O93367 gallus gall
15	9	2.2	309	1 HXA4_CHICK	P17277 gallus gall
16	9	2.2	345	1 CEBB_HUMAN	P17676 homo sapien
17	9	2.2	348	1 CEBB_BOVIN	O02755 bos taurus
18	9	2.2	353	1 CEBB_BOVIN	O02754 bos taurus
19	9	2.2	367	1 P53_CHICK	P10360 gallus gall
20	9	2.2	377	1 PFTA_MOUSE	O61239 mus musculu
21	9	2.2	377	1 PFTA_RAT	O04631 rattus norv
22	9	2.2	389	1 FL_ORYSA	O24175 oryza sativ
23	9	2.2	389	1 NDPP_MOUSE	O03173 mus musculu
24	9	2.2	401	1 HME1_MOUSE	P09065 mus musculu
25	9	2.2	410	1 HXA3_HETFR	O91a21 heterodontu
26	9	2.2	415	1 ACRO_PIG	P08001 sus scrofa
27	9	2.2	421	1 ACRO_HUMAN	P10323 homo sapien
28	9	2.2	431	1 ACRO_RABIT	P48038 oryctolagus
29	9	2.2	431	1 SMA6_CHICK	Q9W734 gallus gall
30	9	2.2	436	1 GDF6_BOVIN	P55106 bos taurus
31	9	2.2	440	1 FXGA_CHICK	Q98937 gallus gall
32	9	2.2	440	1 G3PT_MOUSE	O64467 mus musculu
33	9	2.2	449	1 WT1_HUMAN	P19544 homo sapien

34	9	2.2	449	1 WTL_PIG	O62651 sus scrofa
35	9	2.2	450	1 CYSF_TRYBB	P14658 trypanosoma
36	9	2.2	456	1 FXD1_MOUSE	Q61345 mus musculu
37	9	2.2	474	1 VP61_NPVOP	O10270 orgyia pseu
38	9	2.2	477	1 CAP2_RAT	P52481 rattus norv
39	9	2.2	485	1 SSGP_VOLCA	P21997 volvox cart
40	9	2.2	487	1 EBN2_EBV	P12978 epstein-bar
41	9	2.2	488	1 BTB1_MOUSE	P58544 mus musculu
42	9	2.2	497	1 WAS2_HUMAN	Q9y6w5 homo sapien
43	9	2.2	501	1 WASL_RAT	O08816 rattus norv
44	9	2.2	502	1 WASP_HUMAN	P42768 homo sapien
45	9	2.2	505	1 WASL_BOVIN	Q95107 bos taurus

ALIGNMENTS

RESULT 1	
UL33_RCMVM	
ID	UL33_RCMVM STANDARD; PRT; 387 AA.
AC	O12000;
DT	15-JUL-1998 (Rel. 36, Created)
DT	15-JUL-1998 (Rel. 36, Last sequence update)
DT	16-OCT-2001 (Rel. 40, Last annotation update)
DE	G-protein coupled receptor homolog R33.
OS	Rat cytomegalovirus (strain Maastricht).
OC	Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
OC	Betaherpesvirinae; Muromegalovirus.
OX	NCBI_TaxID=79700;
RN	[1]
RP	SEQUENCE FROM N.A.
RX	MEDLINE=98139136; PubMed=9499096;
RA	Beisser P.S., Vink C., Van Dam J.G., Grauls G., Vanherle S.J.,
RA	Bruggeman C.A.;
RT	"The R33 G protein-coupled receptor gene of rat cytomegalovirus plays
RT	an essential role in the pathogenesis of viral infection.";
RL	J. Virol. 72:2352-2363(1998).
RN	[2]
RP	SEQUENCE FROM N.A.
RX	MEDLINE=20366325; PubMed=10906222;
RA	Vink C., Beuken E., Bruggeman C.A.;
RA	"Complete DNA sequence of the rat cytomegalovirus genome.";
RL	J. Virol. 74:7656-7665(2000).
CC	-1- FUNCTION: PLAYS AN IMPORTANT ROLE IN VIVO, IN PARTICULAR IN THE
CC	DISSEMINATION TO OR REPLICATION IN THE SALIVARY GLAND.
CC	-1- SUBCELLULAR LOCATION: Integral membrane protein.
CC	-1- SIMILARITY: BELONGS TO FAMILY 1 OF G-PROTEIN COUPLED RECEPTORS.
CC	-----
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC	the European Bioinformatics Institute. There are no restrictions on its
CC	use by non-profit institutions as long as its content is in no way
CC	modified and this statement is not removed. Usage by and for commercial
CC	entities requires a license agreement (See <a href="http://www.isb-sib.ch/announce/">http://www.isb-sib.ch/announce/</a>
CC	or send an email to <a href="mailto:license@isb-sib.ch">license@isb-sib.ch</a> ).
CC	-----
DR	EMBL; AF232689; AAC58815.1; -
DR	GCRRDB; GCR_1283; -
DR	InterPro; IPR000276; GPCR_Rhodpsn.
DR	Pfam; PF00001; 7tm_1; 1.
DR	PROSITE; PS00237; G-PROTEIN_RECEP_FL_1; 1.
DR	PROSITE; PS50262; G-PROTEIN_RECEP_FL_2; 1.
DR	G-protein coupled receptor; Transmembrane.
KW	DOMAIN 1 33
FT	TRANSMEM 34 61
FT	DOMAIN 62 71
FT	TRANSMEM 72 94
FT	DOMAIN 95 107
FT	TRANSMEM 108 129
FT	DOMAIN 130 150
FT	TRANSMEM 151 169
FT	DOMAIN 170 204
FT	TRANSMEM 205 224

```

FT DOMAIN 225 244 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 245 268 6 (POTENTIAL).
FT DOMAIN 269 285 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 286 309 7 (POTENTIAL).
FT DOMAIN 310 387 CYTOPLASMIC (POTENTIAL).
FT DOMAIN 375 385 POLY-PRO.
FT CARBOHYD 20 20 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 387 AA; 43159 MW; EF80D3F10344D6AE CRC64;

Query Match 2.5%; Score 10; DB 1; Length 387;
Best Local Similarity 100.0%; Pred. No. 0.17; 0; Indels 0; Gaps 0;
Matches 10; Conservative 0; Mismatches 0;

QY 48 PPPPPPPPPN 57
Db 377 PPPPPPPPPN 386
|||||

RESULT 2
ID YPRO_OWEFU STANDARD; PRT; 141 AA.
AC P21260; P21261;
DT 01-MAY-1991 (Rel. 18, Created)
DT 01-MAY-1991 (Rel. 18, Last sequence update)
DT 01-NOV-1995 (Rel. 32, Last annotation update)
DE Hypothetical proline-rich protein (Fragment).
OS Owenia fusiformis.
OC Eukaryota; Metazoa; Annelida; Polychaeta; Palpata; Canaliculipata;
OC Sabellida; Oweniidae; Owenia.
OX NCBI_TaxID=6347;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=90147742; PubMed=2105723;
RA Bakalaria N., Collet J., Planells R., Thouveny Y., Fontes M.;
RT "Presence in invertebrate genomes of sequences characterized by the
RL Biochem. Biophys. Res. Commun. 166:66-73(1990).
DR PIR; A34043; A34043.
DR PIR; B34043; B34043.
KW Hypothetical protein; DNA-binding.
FT NON_TER 1 1
FT DOMAIN 9 58 POLY-PRO.
FT DNA_BIND 98 116 H-T-H MOTIF (POTENTIAL).
FT NON_TER 141 141
SQ SEQUENCE 141 AA; 15745 MW; B294E884D152BDB9 CRC64;

Query Match 2.2%; Score 9; DB 1; Length 141;
Best Local Similarity 100.0%; Pred. No. 0.58;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 48 PPPPPPPPP 56
Db 9 PPPPPPPPP 17
|||||

RESULT 3
ID Y084_RICPR STANDARD; PRT; 165 AA.
AC Q92B63;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical protein RP084.
GN RP084.
OS Rickettsia prowazekii.
OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
OC Rickettsiaceae; Rickettsiae; Rickettsia.
OX NCBI_TaxID=782;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MADRID E;

```

```

RX MEDLINE=99039499; PubMed=9823893;
RA Andersson S.G.E., Zomorodipour A., Andersson J.O.,
RA Sicheritz-Ponten T., Alsmark U.C.M., Podowski R.M., Naeslund A.K.,
RA Eriksson A.-S., Winkler H.H., Kurland C.G.;
RT "The genome sequence of Rickettsia prowazekii and the origin of
RT mitochondria.";
RL Nature 396:133-140(1998).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AJ235270; CAA14554.1; --
CC Hypothetical protein; Complete proteome.
CC SEQUENCE 165 AA; 18232 MW; E844F1DF89C0A763 CRC64;

Query Match 2.2%; Score 9; DB 1; Length 165;
Best Local Similarity 100.0%; Pred. No. 0.66;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 48 PPPPPPPPP 56
Db 36 PPPPPPPPP 44
|||||

RESULT 4
ID TONB_XANCP STANDARD; PRT; 223 AA.
AC Q34261.
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE TonB protein.
GN TONB.
OS Xanthomonas campestris (pv. campestris).
OC Bacteria; Proteobacteria; gamma subdivision; Xanthomonas group;
OC Xanthomonas.
OX NCBI_TaxID=340;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=98037510; PubMed=9371459;
RA Wiggerich H.G., Klauke B., Koepf R., Priefer U.B., Puehler A.;
RT "Unusual structure of the tonB-exb DNA region of Xanthomonas
RT campestris pv. campestris: tonB, exbB, and exbD1 are essential for
RT ferric iron uptake, but exbD2 is not.";
RL J. Bacteriol. 179:7103-7110(1997).
CC -!- FUNCTION: INTERACTS WITH OUTER MEMBRANE RECEPTOR PROTEINS THAT
CC CARRY OUT HIGH-AFFINITY BINDING AND ENERGY DEPENDENT UPTAKE INTO
CC THE PERIPLASMIC SPACE OF SPECIFIC SUBSTRATES. IT COULD ACT TO
CC TRANSDUCE ENERGY FROM THE CYTOPLASMIC MEMBRANE TO SPECIFIC ENERGY-
CC REQUIRING PROCESSES IN THE OUTER MEMBRANE, RESULTING IN THE
CC RELEASE INTO THE PERIPLASM OF LIGANDS BOUND BY THESE OUTER
CC MEMBRANE PROTEINS (BY SIMILARITY).
CC -!- SUBUNIT: HOMODIMER. FORMS A COMPLEX WITH THE ACCESSORY PROTEINS
CC EXBB AND EXBD (BY SIMILARITY).
CC -!- SUBCELLULAR LOCATION: PERIPLASMIC. ANCHORED TO THE CYTOPLASMIC
CC MEMBRANE VIA ITS N-TERMINAL SIGNAL-LIKE SEQUENCE, SPANS THE
CC PERIPLASM (BY SIMILARITY).
CC -!- SIMILARITY: BELONGS TO THE TONB FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC

```

DR EMBL; Z95386; CAB08610.1; --  
DR InterPro; IPR003538; TonB.  
DR PRINTS; PR01374; TonBPROTEIN.  
KW Transport; Protein transport; Inner membrane; Periplasmic;  
KW Transmembrane; Signal-anchor; Repeat.  
FT DOMAIN 1 23 CYTOPLASMIC (POTENTIAL).  
FT TRANSMEM 24 44 SIGNAL-ANCHOR (POTENTIAL).  
FT DOMAIN 45 223 PERIPLASMIC (POTENTIAL).  
FT DOMAIN 66 78 POLY-PRO.  
FT DOMAIN 82 85 POLY-PRO.  
FT DOMAIN 94 101 POLY-PRO.  
SQ SEQUENCE 223 AA; 23598 MW; D7C6886FDF06FDA9 CRC64;

Query Match 2.2%; Score 9; DB 1; Length 223;  
Best Local Similarity 100.0%; Pred. No. 0.86; Indels 0;  
Matches 9; Conservative 0; Mismatches 0; Gaps 0;

Qy 48 P P P P P P P P P P 56  
| | | | | | | | | |  
Db 66 P P P P P P P P P P 74

RESULT 5  
VE4\_HPV08  
ID VE4\_HPV08 STANDARD; PRT; 229 AA.  
AC P06425;  
DT 01-JAN-1988 (Rel. 06, Created)  
DT 01-JAN-1988 (Rel. 06, Last sequence update)  
DT 01-AUG-1992 (Rel. 23, Last annotation update)  
DE Probable E4 protein.  
OS Human papillomavirus type 8.  
OC Viruses; dsDNA viruses, no RNA stage; Papillomaviridae;  
OC Papillomavirus.  
OX NCBI\_TaxID=10579;  
RN NCBI\_TaxID=10579;  
RP SEQUENCE FROM N.A.  
RX MEDLINE=86200410; PubMed=3009874;  
RA Fuchs P.G., Iftner T., Weninger J., Pfister H.;  
RT "Epidermodyplasia verruciformis-associated human papillomavirus 8:  
RL genomic sequence and comparative analysis.";  
RL J. Virol. 58:626-634(1986).  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL; M12737; --; NOT\_ANNOTATED\_CDS.  
DR PIR; A03677; W4WL8.  
KW Early protein.  
SQ SEQUENCE 229 AA; 24205 MW; 5A5366640FAB45C2 CRC64;

Query Match 2.2%; Score 9; DB 1; Length 229;  
Best Local Similarity 100.0%; Pred. No. 0.88;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 49 P P P P P P P P P P 57  
| | | | | | | | | |  
Db 178 P P P P P P P P P P 186

RESULT 6  
HXB4\_MOUSE  
ID HXB4\_MOUSE STANDARD; PRT; 250 AA.  
AC P10284;  
DT 01-MAR-1989 (Rel. 10, Created)  
DT 01-MAR-1989 (Rel. 10, Last sequence update)  
DT 01-OCT-1996 (Rel. 34, Last annotation update)

DE Homeobox protein Hox-B4 (Hox-2.6).  
GN HOXB4 OR HOXB-4 OR HOX-2.6.  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
RN NCBI\_TaxID=10090;  
[1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=89091992; PubMed=2463210;  
RA Graham A., Papalopulu N., Lorimer J., Mcvey J.H., Tuddenham E.G.D.,  
RA Krumlauf R.;  
RT "Characterization of a murine homeo box gene, Hox-2.6, related to the  
RL Genes Dev. 2:1424-1438(1988).  
CC -|- FUNCTION: SEQUENCE-SPECIFIC TRANSCRIPTION FACTOR WHICH IS PART OF  
CC A DEVELOPMENTAL REGULATORY SYSTEM THAT PROVIDES CELLS WITH  
CC SPECIFIC POSITIONAL IDENTITIES ON THE ANTERIOR-POSTERIOR AXIS.  
CC -|- SUBCELLULAR LOCATION: Nuclear.  
CC -|- SIMILARITY: BELONGS TO THE ANTP FAMILY OF HOMEBOX PROTEINS.  
CC "DEFORMED" SUBFAMILY.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL; M36654; AAA37848.1; --  
DR PIR; A31757; A31757.  
DR HSSP; P02833; 9ANT.  
DR TRANSFAC; T01726; --  
DR MGD; MGI:96185; Hoxb4.  
DR InterPro; IPR001827; Antennapedia.  
DR InterPro; IPR001356; Homeobox.  
DR Pfam; PF00046; homeobox; 1.  
DR PRINTS; PR00025; ANTENNAPEIDIA.  
DR PRINTS; PR00024; HOMEBOX.  
DR SMART; SM00389; HOX; 1.  
DR PROSITE; PS00027; HOMEBOX\_1; 1.  
DR PROSITE; PS00032; ANTENNAPEIDIA; 1.  
DR PROSITE; PS50071; HOMEBOX\_2; 1.  
DR Homeobox; DNA-binding; Developmental protein; Nuclear protein;  
KW Transcription regulation.  
FT DOMAIN 15 138 PRO-RICH; PART OF THE TRANSCRIPTIONAL  
FT DOMAIN 71 86 ACTIVATION DOMAIN.  
FT DOMAIN 140 145 ANTP-TYPE HEXAPEPTIDE.  
FT DNA\_BIND 161 220 HOMEBOX.  
FT SEQUENCE 250 AA; 27519 MW; D09D477A0E585BE6 CRC64;

Query Match 2.2%; Score 9; DB 1; Length 250;  
Best Local Similarity 100.0%; Pred. No. 0.94;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 48 P P P P P P P P P P 56  
| | | | | | | | | |  
Db 73 P P P P P P P P P P 81

RESULT 7  
HXB4\_HUMAN  
ID HXB4\_HUMAN STANDARD; PRT; 251 AA.  
AC P17483; O9NTA0;  
DT 01-AUG-1990 (Rel. 15, Created)  
DT 01-JUN-1994 (Rel. 29, Last sequence update)  
DT 01-MAR-2002 (Rel. 41, Last annotation update)  
DE Homeobox protein Hox-B4 (Hox-2F) (Hox-2.6).  
GN HOXB4 OR HOX2F.  
OS Homo sapiens (Human).  
OS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=91153613; PubMed=1981366;  
 RA Peverali F.A., D'Esposito M., Acampora D., Bunone G., Negri M.,  
 RA Faiella A., Stornaiuolo A., Pannese M., Migliacchio E., Simeone A.,  
 RA Valle G.D., Boncinelli E.;  
 RT "Expression of HOX homeogenes in human neuroblastoma cell culture  
 lines.";  
 RL Differentiation 45:61-69(1990).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RX Kidd K.K., Busyina V., Demille M.M.C., Speed W.C., Ruggeri V.,  
 RA Kidd J.R., Pakstis A.J.;  
 RT "Overall linkage disequilibrium in 33 populations for highly  
 RT informative multistate haplotypes spanning the HOXB gene cluster.";  
 RL Am. J. Hum. Genet. 67:235-235(2000).  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=20538492; PubMed=11085749;  
 RA Giannola D.M., Shlomchik W.D., Jegathesan M., Liebowitz D.,  
 RA Abrams C.S., Kadesch T., Dancis A., Emerson S.G.;  
 RT "Hematopoietic expression of HOXB4 is regulated in normal and leukemic  
 RT stem cells through transcriptional activation of the HOXB4 promoter by  
 RT upstream stimulating factor (USF)-1 and USF-2.";  
 RL J. Exp. Med. 192:1479-1490(2000).  
 RN [4]  
 RP SEQUENCE OF 6-251 FROM N.A.  
 RX TISSUE-Testis;  
 RA Bloecker H., Boecher M., Brandt P., Mewes H.-W., Gassenhuber J.,  
 RA Wiemann S.;  
 RL Submitted (JAN-2000) to the EMBL/GenBank/DBJ databases.  
 RN [5]  
 RP SEQUENCE OF 160-227 FROM N.A.  
 RX TISSUE-Placenta;  
 RA MEDLINE=89378558; PubMed=2570724;  
 RA Giampolo A., Acampora D., Zappavigna V., Pannese M.,  
 RA D'Esposito M., Care A., Faiella A., Stornaiuolo A., Russo G.,  
 RA Simeone A., Boncinelli E., Peschle C.;  
 RT "Differential expression of human HOX-2 genes along the anterior-  
 RT posterior axis in embryonic central nervous system.";  
 RL Differentiation 40:191-197(1989).  
 RN [6]  
 RP SEQUENCE OF 162-227 FROM N.A.  
 RX MEDLINE=90215256; PubMed=2576652;  
 RA Boncinelli E., Acampora D., Pannese M., D'Esposito M., Somma R.,  
 RA Gaudino G., Stornaiuolo A., Caffiero M., Faiella A., Simeone A.;  
 RT "Organization of human class I homeobox genes.";  
 RL Genome 31:745-756(1989).  
 CC -1- FUNCTION: SEQUENCE-SPECIFIC TRANSCRIPTION FACTOR WHICH IS PART OF  
 CC A DEVELOPMENTAL REGULATORY SYSTEM THAT PROVIDES CELLS WITH  
 CC SPECIFIC POSITIONAL IDENTITIES ON THE ANTERIOR-POSTERIOR AXIS.  
 CC -1- SUBCELLULAR LOCATION: Nuclear.  
 CC -1- DEVELOPMENTAL STAGE: EXPRESSED IN WHOLE EMBRYOS AND FETUSES AT  
 CC 5-9 WEEKS FROM CONCEPTION.  
 CC -1- SIMILARITY: BELONGS TO THE ANTP FAMILY OF HOMEBOX PROTEINS.  
 CC "DEFORMED" SUBFAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; X16174; CAA34296.1;  
 DR EMBL; AF287967; AAG31554.1;  
 DR EMBL; AF307160; AAG45052.1;  
 DR EMBL; AL137449; CAB70742.1;  
 DR PIR; B60492; B60492.  
 DR PIR; C37042; C37042.

DR PIR; S15543; S15543.  
 DR HSP; P02833; 9ANT.  
 DR TRANSFAC; T01727; -.  
 DR MIM; 142965; -.  
 DR InterPro; IPR001827; Antennapedia.  
 DR InterPro; IPR001356; Homeobox.  
 DR Pfam; PF00046; homeobox; 1.  
 DR PRINTS; PR00025; ANTENNAPEIDIA.  
 DR PRINTS; PR00024; HOMEBOX.  
 DR SMART; SM00389; HOX; 1.  
 DR PROSITE; PS00027; HOMEBOX\_1; 1.  
 DR PROSITE; PS00032; ANTENNAPEIDIA; 1.  
 DR PROSITE; PS00071; HOMEBOX\_2; 1.  
 KW Homeobox; DNA-binding; Developmental protein; Nuclear protein;  
 KW Transcription regulation.  
 FT DOMAIN 15 139 PRO-RICH; PART OF THE TRANSCRIPTIONAL  
 FT DOMAIN 71 87 ACTIVATION DOMAIN.  
 FT DOMAIN 141 146 POLY-PRO.  
 FT DNA\_BIND 162 221 ANTP-TYPE HEXAPEPTIDE.  
 FT HOMEBOX 251 AA; 27604 MW; 23EEID0DDCCE2DB4 CRC64;  
 SQ SEQUENCE 2.2%; Score 9; DB 1; Length 251;  
 Query Match Best Local Similarity 100.0%; Pred. No. 0.95;  
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 48 P P P P P P P P P P 56  
 DB 73 P P P P P P P P P P 81  
 RESULT 8  
 TRPC\_MYCTU STANDARD; PRT; 272 AA.  
 ID TRPC\_MYCTU  
 AC 06129;  
 DT 30-MAY-2000 (Rel. 39, Created)  
 DT 30-MAY-2000 (Rel. 39, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Indole-3-glycerol phosphate synthase (EC 4.1.1.48) (IGPS).  
 GN TRPC OR RV1611 OR MT1646 OR MTCV01B2.03.  
 OS Mycobacterium tuberculosis.  
 OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae; Mycobacterium.  
 OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.  
 OX NCBI\_TaxID=1773;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=H37RV;  
 RX MEDLINE=98295987; PubMed=9634230;  
 RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,  
 RA Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekala F.,  
 RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,  
 RA Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,  
 RA Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,  
 RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,  
 RA Rutter S., Seeger K., Skelton S., Squares S., Squares R.,  
 RA Stulston J.E., Taylor K., Whitehead S., Barrell B.G.;  
 RT "Deciphering the biology of Mycobacterium tuberculosis from the  
 RT complete genome sequence.";  
 RL Nature 393:537-544(1998).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=CDC 1551 / Oshkosh;  
 RA Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,  
 RA Peterson J., DeBoy R., Dodson R., Gwinn M.L., Haft D., Hickey E.,  
 RA Kolonay J.F., Nelson W.C., Umayam L.A., Esmolaeva M.D., Salzberg S.L.,  
 RA Delcher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A.,  
 RA Bishai W.;  
 RT "Whole genome comparison of Mycobacterium tuberculosis clinical and  
 RT laboratory strains.";  
 RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.  
 CC -1- CATALYTIC ACTIVITY: 1-(2-carboxyphenylamino)-l-deoxy-D-ribose 5-  
 CC phosphate = 1-(indol-3-yl)glycerol 3-phosphate + CO(2) + H(2)O.



CC -|- PATHWAY: FOURTH STEP IN BIOSYNTHESIS OF TRYPTOPHAN.  
 CC -|- SIMILARITY: BELONGS TO THE TRPC FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL; Z95554; CAB08905.1; -;  
 CC EMBL; AB007029; AKA43915.1; ALT\_INT.  
 CC HSSP; P00909; IPII.  
 CC TIGR; MT1646; -;  
 CC Tuberculin; RV1611; -;  
 CC InterPro; IPR003009; FMN\_enzyme.  
 CC InterPro; IPR001468; IGPS.  
 CC Pfam; PF00218; IGPS; 1.  
 CC ProDom; PD001511; IGPS; 1.  
 CC PROSITE; PS00614; IGPS; 1.  
 CC Tryptophan biosynthesis; Lyase; Decarboxylase; Complete proteome.  
 KW Tryptophan biosynthesis; Lyase; Decarboxylase; Complete proteome.  
 FT DOMAIN 32 37 POLY-ALA.  
 FT CONFLICT 164 164 T -> I (IN REF. 2).  
 SQ SEQUENCE 272 AA; 28023 MW; 9CA29D0F0FAC76C2 CRC64;

Query Match 2.2%; Score 9; DB 1; Length 272;  
 Best Local Similarity 100.0%; Pred. No. 1;  
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 140 AAAAAAPPP 148  
 Db 32 AAAAAAPPP 40

RESULT 9  
 ID FASL\_HUMAN STANDARD; PRT; 281 AA.  
 AC P48023;  
 DT 01-FEB-1996 (Rel. 33, Created)  
 DT 01-FEB-1996 (Rel. 33, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE FAS antigen ligand (Apoptosis antigen ligand) (APTL) (CD178 antigen).  
 GN TNFSF6 OR APTL1G1 OR FASL.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=95105731; PubMed=7528780;  
 RA Alderson M.;  
 RT "Fas ligand mediates activation-induced cell death in human T  
 RT lymphocytes.";  
 RL J. Exp. Med. 181:71-77(1995).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=95127560; PubMed=7826947;  
 RA Takahashi T., Tanaka M., Inazawa J., Abe T., Suda T., Nagata S.;  
 RT "Human Fas ligand: gene structure, chromosomal location and species  
 RT specificity.";  
 RL Int. Immunol. 6:1567-1574(1994).  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RA Schatzlein C.E., Poehlmann R., Philippson P., Eibel H.;  
 RL Submitted (JUN-1995) to the EMBL/GenBank/DBJ databases.  
 RN [4]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=95071350; PubMed=7980502;  
 RA Mita E., Hayashi N., Iio S., Takehara T., Hijioaka T., Kasahara A.,  
 RA Fusamoto H., Kamada T.;  
 RT "Role of Fas ligand in apoptosis Induced by hepatitis C virus

RT Infection.";  
 RL Biochem. Biophys. Res. Commun. 204:468-474(1994).  
 RN [5]  
 RP SEQUENCE FROM N.A.  
 RA Wilkinson J.;  
 RL Submitted (MAY-1997) to the EMBL/GenBank/DBJ databases.  
 RN [6]  
 RP SEQUENCE OF 1-10 FROM N.A.  
 RC TISSUE=Blood;  
 RA Matsumura M., Nakanishi Y., Ohba Y.;  
 RL Submitted (APR-1998) to the EMBL/GenBank/DBJ databases.  
 CC -|- FUNCTION: CYTOKINE THAT BINDS TO FAS ANTIGEN, A RECEPTOR THAT  
 CC TRANSDUCES THE APOPTOTIC SIGNAL INTO CELLS. MAY BE INVOLVED IN  
 CC CYTOTOXIC T CELL MEDIATED APOPTOSIS AND IN T CELL DEVELOPMENT.  
 CC FAS-ANTIGEN MEDIATED APOPTOSIS MAY HAVE A ROLE IN THE INDUCTION OF  
 CC PERIPHERAL TOLERANCE, IN THE ANTIGEN-STIMULATED SUICIDE OF MATURE  
 CC T CELLS, OR BOTH.  
 CC -|- SUBUNIT: HOMOTRIMER (PROBABLE).  
 CC -|- SUBCELLULAR LOCATION: TYPE II MEMBRANE PROTEIN. MAY BE RELEASED  
 CC INTO THE EXTRACELLULAR FLUID, PROBABLY BY CLEAVAGE FORM THE CELL  
 CC SURFACE.  
 CC -|- SIMILARITY: BELONGS TO THE TUMOR NECROSIS FACTOR FAMILY.  
 CC -|- DATABASE: NAME=PRO; NOTE=CD guide CD178 entry;  
 CC WWW="http://www.ncbi.nlm.nih.gov/prov/guide/338769674\_g.htm".  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL; X89102; CAA61474.1; -;  
 CC EMBL; U08137; AAC50071.1; -;  
 CC EMBL; U11821; AAC50124.1; -;  
 CC EMBL; D38122; BAA07320.1; -;  
 CC EMBL; Z96050; CAB09424.1; -;  
 CC EMBL; AB013303; BAA32542.1; -;  
 CC HSSP; P06804; 2TNF.  
 CC TM: 134638; -;  
 DR InterPro; IPR003263; TNF\_5.  
 DR InterPro; IPR003636; TNF\_abd.  
 DR InterPro; IPR000478; TNF\_family.  
 DR Pfam; PF00229; TNF; 1.  
 DR ProDom; PD002012; TNF\_abd; 1.  
 DR ProDom; PD008600; TNF\_5; 1.  
 DR SMART; SM00207; TNF; 1.  
 DR PROSITE; PS00251; TNF\_1; 1.  
 DR PROSITE; PS00049; TNF\_2; 1.  
 DR Cytokine; Transmembrane; Glycoprotein; Signal-anchor; Apoptosis;  
 KW Antigen.  
 FT DOMAIN 1 80 CYTOPLASMIC (POTENTIAL).  
 FT TRANSMEM 81 102 SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN).  
 FT DOMAIN 103 281 EXTRACELLULAR (POTENTIAL).  
 FT DOMAIN 4 70 PRO-RICH.  
 FT DOMAIN 45 65 POLY-PRO.  
 FT DISULFID 202 233 BY SIMILARITY.  
 FT CARBOHYD 184 184 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 250 250 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 260 260 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 SQ SEQUENCE 281 AA; 31485 MW; ABA6EB358246E9BB CRC64;

Query Match 2.2%; Score 9; DB 1; Length 281;  
 Best Local Similarity 100.0%; Pred. No. 1;  
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 48 pppppppppp 56  
 Db 45 pppppppppp 53



FT DOMAIN 191 197 POLY-PRO.  
FT DOMAIN 213 222 POLY-PRO.  
FT DOMAIN 239 243 POLY-PRO.  
FT CONFLICT 8 MISSING (IN REF. 2).  
FT CONFLICT 65 K -> E (IN REF. 2).  
FT CONFLICT 211 S -> N (IN REF. 2).  
SQ SEQUENCE 289 AA; 31193 MW; A8236F80791CE52B CRC64;

Query Match 2.2%; Score 9; DB 1; Length 289;  
Best Local Similarity 100.0%; Pred. No. 1.1;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 48 PPPPPPPPP 56  
Db 213 PPPPPPPPP 221

RESULT 12  
HSD8\_HUMAN STANDARD; PRT; 290 AA.  
AC P13378;  
DT 01-JAN-1990 (Rel. 13, Created)  
DT 16-OCT-2001 (Rel. 40, Last sequence update)  
DT 01-MAR-2002 (Rel. 41, Last annotation update)  
DE Homeobox protein Hox-D8 (Hox-4E) (Hox-5.4).  
GN HOXD8 OR HOX4E.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Birren B., Linton L., Nusbaum C., Lander E.;  
RL Submitted (AUG-1999) to the EMBL/GenBank/DBJ databases.  
RN [2]  
RP SEQUENCE FROM N.A.  
RA Kosaki K., Kosaki R., Suzuki T., Yoshihashi H., Sasaki K., Matsuo N.;  
RT "A complete mutation analysis panel of human HOX genes.";  
RL Submitted (NOV-2000) to the EMBL/GenBank/DBJ databases.  
RN [3]  
RP SEQUENCE OF 193-287 FROM N.A.  
RX MEDLINE=89306602; PubMed=2568311;  
RA Oliver G., Sidell N., Fiske N., Heinzmann C., Mohandas T.,  
RA Sparkes R.S., de Robertis E.M.;  
RT "Complementary homeo protein gradients in developing limb buds.";  
RL Genes Dev. 3:641-650(1989).  
CC -!- FUNCTION: SEQUENCE-SPECIFIC TRANSCRIPTION FACTOR WHICH IS PART OF  
CC A DEVELOPMENTAL REGULATORY SYSTEM THAT PROVIDES CELLS WITH  
CC SPECIFIC POSITIONAL IDENTITIES ON THE ANTERIOR-POSTERIOR AXIS.  
CC -!- SUBCELLULAR LOCATION: Nuclear.  
CC -!- SIMILARITY: BELONGS TO THE ANTP FAMILY OF HOMEBOX PROTEINS.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL; AC009336; -; NOT\_ANNOTATED\_CDS.  
DR EMBL; AY014304; AAG42152.1; -  
DR EMBL; AY014303; AAG42152.1; JOINED.  
DR EMBL; X15507; CAA33529.1; -  
DR PIR; B32830; B32830.  
DR PIR; S05957; S05957.  
DR HSP; P02833; 9ANT.  
DR TRANSFAC; T03332; -  
DR MIM; 142985; -  
DR InterPro; IPR001827; Antennapedia.  
DR InterPro; IPR001356; Homeobox.  
DR Pfam; PF00046; homeobox; 2.

DR PRINTS; PRO0025; ANTENNAPEDIA.  
DR PRINTS; PRO0024; HOMEBOX.  
DR SMART; SM00389; HOX; 1.  
DR PROSITE; PS00027; HOMEBOX\_1; 1. PARTIAL.  
DR PROSITE; PS00032; ANTENNAPEDIA; 1.  
DR PROSITE; PS0071; HOMEBOX\_2; 1.  
KW Homeobox; DNA-binding; Developmental protein; Nuclear protein;  
KW Transcription regulation.  
FT DOMAIN 15 23 POLY-ALA.  
FT DOMAIN 45 50 POLY-ALA.  
FT DOMAIN 109 123 POLY-PRO.  
FT DNA\_BIND 197 256 HOMEBOX.  
FT CONFLICT 287 G -> A (IN REF. 3).  
SQ SEQUENCE 290 AA; 31910 MW; 75FF95A73E2AA85F CRC64;

Query Match 2.2%; Score 9; DB 1; Length 290;  
Best Local Similarity 100.0%; Pred. No. 1.1;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 48 PPPPPPPPP 56  
Db 115 PPPPPPPPP 123

RESULT 13  
SMN\_HUMAN STANDARD; PRT; 294 AA.  
ID SMN\_HUMAN  
AC Q16637; Q13119;  
DT 01-NOV-1997 (Rel. 35, Created)  
DT 01-NOV-1997 (Rel. 35, Last sequence update)  
DT 01-MAR-2002 (Rel. 41, Last annotation update)  
DE Survival motor neuron protein (Component of gems 1) (Gemin1).  
GN (SMN1 OR SMN OR SMNT) AND (SMN2 OR SMNC).  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A., ALTERNATIVE SPLICING, AND VARIANT SMA CYS-272.  
RC TISSUE=Fetal brain;  
RX MEDLINE=95112343; PubMed=7813012;  
RA Lefebvre S., Buerklen L., Reboullet S., Clermont O., Bulet P.,  
RA Viollet L., Benichou B., Cruaud C., Millasseau P., Zeviani M.,  
RA le Paslier D., Frezal J., Cohen D., Weissenbach J., Munnich A.,  
RA Melki J.;  
RT "Identification and characterization of a spinal muscular atrophy-  
RT determining gene.";  
RL Cell 80:155-165(1995).  
RN [2]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=96435930; PubMed=8838816;  
RA Buerklen L., Lefebvre S., Clermont O., Bulet P., Viollet L.,  
RA Cruaud C., Munnich A., Melki J.;  
RT "Structure and organization of the human survival motor neurone (SMN)  
RT gene.";  
RL Genomics 32:479-482(1996).  
RN [3]  
RP SEQUENCE FROM N.A.  
RA Chen Q., Baird S.D., Besner-Johnston A., Farahani R., Xuan J.Y.,  
RA Kang X., Lefebvre C., Korneluk R.G., Mackenzie A.E.;  
RL Submitted (DEC-1996) to the EMBL/GenBank/DBJ databases.  
RN [4]  
RP SEQUENCE FROM N.A.  
RA Jones K., Graves T., McPherson J.;  
RL Submitted (JUN-1998) to the EMBL/GenBank/DBJ databases.  
RN [5]  
RP SEQUENCE FROM N.A., AND ALTERNATIVE SPLICING.  
RC TISSUE=Skeletal muscle;  
RX MEDLINE=95367018; PubMed=7639755;  
RA Gennarelli M., Lucarelli M., Capon F., Pizzuti A., Merlini L.,  
RA Angelini C., Novelli G., Dallapiccola B.;  
RT "Survival motor neuron gene transcript analysis in muscles from spinal

RT muscular atrophy patients.";  
RL Biochem. Biophys. Res. Commun. 213:342-348(1995).  
RN [6]  
RX SEQUENCE OF 2-294 FROM N.A.  
RP MEDLINE=96018066; PubMed=7552146;  
RA van der Steege G., Draaijers T.G., Grootsholten P.M., Osinga J.,  
RA Anzevino R., Velona I., Den Dunnen J.T., Scheffer H., Brahe C.,  
RA van Ommen G.J.,  
RT "A provisional transcript map of the spinal muscular atrophy (SMA)  
RT critical region.";  
RL Eur. J. Hum. Genet. 3:87-95(1995).  
RN [7]  
RP SUBCELLULAR LOCATION  
RX MEDLINE=96312953; PubMed=8670859;  
RA Liu Q., Dreyfuss G.;  
RT "A novel nuclear structure containing the survival of motor neurons  
RT protein.";  
RL EMBO J. 15:3555-3565(1996).  
RN [8]  
RP FUNCTION.  
RX MEDLINE=99059497; PubMed=9845364;  
RA Pellizzoni L., Kataoka N., Charroux B., Dreyfuss G.;  
RT "A novel function for SMN, the spinal muscular atrophy disease gene  
RT product, in pre-mRNA splicing.";  
RL Cell 95:615-624(1998).  
RN [9]  
RP TISSUE SPECIFICITY.  
RX MEDLINE=97402201; PubMed=9259265;  
RA Coovert D.D., Le T.T., McAndrew P.E., Strasswimmer J., Crawford T.O.,  
RA Mendell J.R., Coulson S.E., Androphy E.J., Prior T.W., Burghes A.H.;  
RT "The survival motor neuron protein in spinal muscular atrophy.";  
RL Hum. Mol. Genet. 6:1205-1214(1997).  
RN [10]  
RP INTERACTION WITH SIP1.  
RX MEDLINE=97462902; PubMed=9323129;  
RA Liu Q., Fischer U., Wang F., Dreyfuss G.;  
RT "The spinal muscular atrophy disease gene product, SMN, and its  
RT associated protein SIP1 are in a complex with spliceosomal snRNP  
RT proteins.";  
RL Cell 90:1013-1021(1997).  
RN [11]  
RP STRUCTURE BY NMR OF 82-169.  
RX MEDLINE=20577270; PubMed=11135666;  
RA Selenko P., Sprangers R., Stier G., Buhler D., Fischer U., Sattler M.;  
RT "SMN tudor domain structure and its interaction with the Sm  
RT proteins.";  
RL Nat. Struct. Biol. 8:27-31(2001).  
RN [12]  
RP VARIANT SMA VAL-279.  
RX MEDLINE=97227297; PubMed=9147655;  
RA Talbot K., Ponting C.P., Theodosiou A.M., Rodrigues N.R., Surtees R.,  
RA Mountford R., Davies K.E.;  
RT "Missense mutation clustering in the survival motor neuron gene: a  
RT role for a conserved tyrosine and glycine rich region of the protein  
RT in RNA metabolism?";  
RL Hum. Mol. Genet. 6:497-500(1997).  
RN [13]  
RP VARIANTS SMA ILE-262 AND ILE-274.  
RX MEDLINE=97301781; PubMed=9158159;  
RA Hahnen E., Schoenling J., Rudnik-Schoeneborn S., Raschke H.,  
RA Zerres K., Wirth B.;  
RT "Missense mutations in exon 6 of the survival motor neuron gene in  
RT patients with spinal muscular atrophy (SMA).";  
RL Hum. Mol. Genet. 6:821-825(1997).  
RN [14]  
RP VARIANTS SMA LEU-245 AND CYS-272.  
RX MEDLINE=20194955; PubMed=10732817;  
RA Rochette C.F., Surh L.C., Ray P.N., McAndrew P.E., Prior T.W.,  
RA Burghes A.H.M., Vanasse M., Smard L.R.;  
RT "Molecular diagnosis of non-deletion SMA patients using quantitative  
RT PCR of SMN exon 7.";  
RL Neurogenetics 1:141-147(1997).  
RN [15]  
RP VARIANT SMA GLY-2.  
RX MEDLINE=99057509; PubMed=9837824;  
RA Parsons D.W., McAndrew P.E., Iannaccone S.T., Mendell J.R.,  
RA Burghes A.H., Prior T.W.;  
RT "Intragenic telSMN mutations: frequency, distribution, evidence of a  
RT founder effect, and modification of the spinal muscular atrophy  
RT phenotype by cenSMN copy number.";  
RL Am. J. Hum. Genet. 63:1712-1723(1998).  
RN [16]  
RP VARIANT SMA CYS-279.  
RX MEDLINE=20194940; PubMed=10732802;  
RA Wang C.H., Papendick B.D., Bruinsma P., Day J.K.;  
RT "Identification of a novel missense mutation of the smnt gene in two  
RT siblings with spinal muscular atrophy.";  
RL Neurogenetics 1:273-276(1998).  
CC -!- FUNCTION: THE SMN COMPLEX PLAYS AN ESSENTIAL ROLE IN SPLICEOSOMAL  
CC SNRP ASSEMBLY IN THE CYTOPLASM, AND IS REQUIRED FOR PRE-MRNA  
CC SPLICING IN THE NUCLEUS.  
CC -!- SUBUNIT: FORMS A STABLE HETEROMERIC COMPLEX WITH SURVIVAL OF MOTOR  
CC NEURON PROTEIN INTERACTING PROTEIN 1 (SIP1/GEMIN2), GEMIN3 AND  
CC GEMIN4. THIS COMPLEX IS THOUGHT TO PLAY AN ESSENTIAL ROLE IN  
CC BIOGENESIS OF THE SMALL NUCLEAR RIBONUCLEOPROTEIN PARTICLES  
CC (SNRNPs). INTERACTS WITH FIBRILLARIN AND WITH SEVERAL SPLICEOSOMAL  
CC SNRP CORE SM PROTEINS, INCLUDING B/B', SM D AND SM E PROTEINS AND  
CC WITH ILF3.  
CC -!- SUBCELLULAR LOCATION: CYTOPLASMIC AND NUCLEAR. LOCALIZED IN  
CC SUBNUCLEAR STRUCTURES NEXT TO COILED BODIES, CALLED GEMs, WHICH  
CC ARE HIGHLY ENRICHED IN SPLICEOSOMAL SNRNPs.  
CC -!- ALTERNATIVE PRODUCTS: 4 ISOFORMS: SMN/FULL-SMN (SHOWN HERE), SMN-  
CC DELTA5/ISO5-SMN, SMN-DELTA7/ISO7-SMN AND SMN-DELTA57/ISO57-SMN;  
CC SEEM TO BE PRODUCED BY ALTERNATIVE SPLICING. THE SMN-DELTA7  
CC ISOFORM IS THOUGHT TO BE A NONFUNCTIONAL PROTEIN THAT LACKS THE  
CC CAPACITY TO OLIGOMERIZE AND, THUS CANNOT INTERACT WITH SM  
CC PROTEINS. SMN1 PRIMARILY PRODUCES SMN WHEREAS SMN2 EXPRESSES  
CC ABUNDANT LEVELS OF SMN-DELTA7.  
CC -!- TISSUE SPECIFICITY: EXPRESSED IN A WIDE VARIETY OF TISSUES.  
CC EXPRESSED AT HIGH LEVELS IN BRAIN, KIDNEY AND LIVER, MODERATE  
CC LEVELS IN SKELETAL AND CARDIAC MUSCLE, AND LOW LEVELS IN  
CC FIBROBLASTS AND LYMPHOCYTES. ALSO SEEN AT HIGH LEVELS IN SPINAL  
CC CORD.  
CC -!- DISEASE: MUTATED FORMS OF SMN HAVE BEEN FOUND IN INDIVIDUALS WITH  
CC SPINAL MUSCULAR ATROPHIES (SMA). SMAS ARE FATAL AUTOSOMAL  
CC RECESSIVE DISORDERS SUBCLASSIFIED AS TYPE I (WERDNIG-HOFFMANN  
CC DISEASE), TYPE II (INTERMEDIATE FORM), AND TYPE III (WOLFFHART-  
CC KUGELBERG-WELANDER DISEASE) BASED UPON THE AGE OF ONSET AND  
CC CLINICAL SEVERITY. THESE NEURODEGENERATIVE DISORDERS ARE  
CC CHARACTERIZED BY DEGENERATION OF LOWER MOTOR NEURONS, LEADING TO  
CC PROGRESSIVE PARALYSIS MUSCULAR ATROPHY. CONCERNS 1 IN 6000  
CC NEWBORNS  
CC -!- MISCELLANEOUS: THE SMN GENE IS PRESENT IN TWO HIGHLY HOMOLOGOUS  
CC AND FUNCTIONAL COPIES (TELSMN AND CENSMN). THE TELOMERIC COPY OF  
CC SMN GENE SEEMS TO BE THE SMA-DETERMINING GENE WHILE THE  
CC CENTROMERIC COPY SEEMS UNAFFECTED.  
CC -!- SIMILARITY: BELONGS TO THE SMN FAMILY.  
CC -!- SIMILARITY: CONTAINS 1 TUDOR DOMAIN.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
CC EMBL; U43883; AAC50473.1; -  
DR EMBL; U43876; AAC50473.1; JOINED.  
DR EMBL; U43877; AAC50473.1; JOINED.  
DR EMBL; U43878; AAC50473.1; JOINED.  
DR EMBL; U43880; AAC50473.1; JOINED.  
DR EMBL; U43881; AAC50473.1; JOINED.  
DR EMBL; U43882; AAC50473.1; JOINED.  
DR EMBL; U18423; AAC66242.1; -  
DR EMBL; U80017; AAC52048.1; -

DR EMBL; AC005031; AAC62262.1; -;  
DR EMBL; U21914; AAA64505.1; -;  
DR MIM; 600354; -;  
DR MIM; 601627; -;  
DR MIM; 253300; -;  
DR MIM; 253400; -;  
DR MIM; 253550; -;  
DR PDB; 1G5V; 02-MAY-01.

Query Match 2.2%; Score 9; DB 1; Length 294;  
Best Local Similarity 100.0%; Pred. No. 1.1;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 48 PPPPPPPPP 56  
| | | | | | | | | |  
Db 217 PPPPPPPPP 225

RESULT 14  
HX1M\_CHICK STANDARD; PRT; 297 AA.  
AC O93367;  
DT 15-JUL-1999 (Rel. 38, Created)  
DT 15-JUL-1999 (Rel. 38, Last sequence update)  
DT 15-JUL-1999 (Rel. 38, Last annotation update)  
DE Homeobox protein Hox-11L2 (Homeobox TLX-3).  
GN HOX11L2 OR TLX3.  
OS Gallus gallus (Chicken).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae;  
OC Gallus.  
OX NCBI\_TaxID=9031;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-RHODE ISLAND RED;  
RX MEDLINE=98315212; PubMed=9651221;  
RA Logan C.C., Wingate R.J.T., McKay I.J., Lumsden A.;  
RT "Hlx-1 and Hlx-3 homeobox gene expression in cranial sensory ganglia  
RT and hindbrain of the chick embryo: markers of patterned  
RT connectivity".  
RL J. Neurosci. 18:5389-5402(1998).  
CC -1- FUNCTION: SEEMS TO BE INVOLVED IN THE DEVELOPMENT OF CRANIAL  
CC SENSORY INNERVATION FROM PERIPHERAL GANGLIA.  
CC -1- SUBCELLULAR LOCATION: Nuclear (By similarity).  
CC -1- TISSUE SPECIFICITY: EXPRESSION IS RESTRICTED TO NEURONES IN THE  
CC PERIPHERAL AND CENTRAL NERVOUS SYSTEM.

-----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
-----

DR EMBL; AF071875; AAC23901.1; -;  
DR HSSP; P14653; 1b72.  
DR InterPro; IPR001827; Antennapedia.  
DR InterPro; IPR001356; Homeobox.  
DR Pfam; PF00046; homeobox; 1.  
DR PRINTS; PR00025; ANTENNAPEDIA.  
DR PRINTS; PR00024; HOMEBOX.  
DR SMART; SM00389; HOX; 1.  
DR PROSITE; PS00027; HOMEBOX\_1; 1.  
DR PROSITE; PS00071; HOMEBOX\_2; 1.  
KW Homeobox; DNA-binding; Developmental protein; Developmental protein.  
FT DOMAIN 33 50 POLY-PRO.  
FT DOMAIN 147 150 POLY-ALA.  
FT DNA\_BIND 172 231 HOMEBOX.  
SQ SEQUENCE 297 AA; 32369 MW; 0C7C9A73E371A38E CRC64;

Query Match

2.2%; Score 9; DB 1; Length 297;

Best Local Similarity 100.0%; Pred. No. 1.1;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Qy 48 PPPPPPPPP 56  
| | | | | | | | | |  
Db 33 PPPPPPPPP 41

RESULT 15

HXA4\_CHICK STANDARD; PRT; 309 AA.  
ID HXA4\_CHICK  
AC P17277;  
DT 01-AUG-1990 (Rel. 15, Created)  
DT 01-AUG-1990 (Rel. 15, Last sequence update)  
DT 01-JUL-1993 (Rel. 26, Last annotation update)  
DE Homeobox protein Hox-A4 (Chox-1.4).  
GN HXA4 OR CHOX-1.4.  
OS Gallus gallus (Chicken).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae;  
OC Gallus.  
OX NCBI\_TaxID=9031;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=90245562; PubMed=1970866;  
RA Sasaki H., Yokoyama E., Kuroiwa A.;  
RT "Specific DNA binding of the two chicken Deformed family homeodomain  
RT proteins, Chox-1.4 and Chox-a".  
RL Nucleic Acids Res. 18:1739-1747(1990).  
RN [2]  
RP SEQUENCE OF 207-273 FROM N.A.  
RC STRAIN-COMET HUBBARD HYBRID;  
RX MEDLINE=90326535; PubMed=1973835;  
RA Scotting P.J., Hewitt M., Keynes R.J.;  
RT "Isolation and analysis of chick homeobox cDNA clones".  
RL Nucleic Acids Res. 18:3999-3999(1990).  
CC -1- FUNCTION: SEQUENCE-SPECIFIC TRANSCRIPTION FACTOR WHICH IS PART OF  
CC A DEVELOPMENTAL REGULATORY SYSTEM ON THE ANTERIOR-POSTERIOR AXIS.  
CC SPECIFIC POSITIONAL IDENTITIES ON THE 5'-FLANKING SEQUENCE OF ITS CODING  
CC REGION: BINDS TO SITES IN THE 5'-FLANKING SEQUENCE OF THE  
CC HIGH AND LOW AFFINITY BINDING SITES ARE TAATGA(C/G) AND CTAATTTT.  
CC -1- SUBCELLULAR LOCATION: Nuclear.  
CC -1- MISCELLANEOUS: THE PROLINE STRETCH WORKS AS A PART OF THE  
CC TRANSCRIPTIONAL ACTIVATION DOMAIN.  
CC -1- SIMILARITY: BELONGS TO THE ANTP FAMILY OF HOMEBOX PROTEINS.  
CC "DEFORMED" SUBFAMILY.

-----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
-----

DR EMBL; X52670; CAA36896.1; -;  
DR EMBL; X52747; CAB57949.1; -;  
DR PIR; S09257; S09257.  
DR PIR; S10883; S10883.  
DR HSSP; P02833; 9ANT.  
DR TRANSFAC; T00128; -;  
DR InterPro; IPR001827; Antennapedia.  
DR InterPro; IPR001356; Homeobox.  
DR Pfam; PF00046; homeobox; 1.  
DR PRINTS; PR00025; ANTENNAPEDIA.  
DR PRINTS; PR00024; HOMEBOX.  
DR SMART; SM00389; HOX; 1.  
DR PROSITE; PS00027; HOMEBOX\_1; 1.  
DR PROSITE; PS00071; HOMEBOX\_2; 1.  
DR PROSITE; PS00032; ANTENNAPEDIA; 1.  
KW Homeobox; DNA-binding; Developmental protein; Nuclear protein;  
KW Transcription regulation.





QY 120 LREQVEPTPEDEDDDDIELRGAAAAAAPPPIIEECPCEDLPKFDGNDPDLAPFMAQCOIF 179  
 D 120 LREQVEPTPEDEDDDDIELRGAAAAAAPPPIIEECPCEDLPKFDGNDPDLAPFMAQCOIF 179  
 Db 61 LREQVEPTPEDEDDDDIELRGAAAAAAPPPIIEECPCEDLPKFDGNDPDLAPFMAQCOIF 120  
 QY 180 MEKTRDFSVDRVRVCFVTSMTGTGAARWASAKLERSHYLMHNYPAFMEMKMHVFEDPQR 239  
 D 180 MEKTRDFSVDRVRVCFVTSMTGTGAARWASAKLERSHYLMHNYPAFMEMKMHVFEDPQR 239  
 Db 121 MEKTRDFSVDRVRVCFVTSMTGTGAARWASAKLERSHYLMHNYPAFMEMKMHVFEDPQR 180  
 QY 240 REVAKRKIRRLROGMGVIDYSNPFQMTAQDLDWNEPALIDQYHEGLSDHITQEEELSHLEV 299  
 D 240 REVAKRKIRRLROGMGVIDYSNPFQMTAQDLDWNEPALIDQYHEGLSDHITQEEELSHLEV 299  
 Db 181 REVAKRKIRRLROGMGVIDYSNPFQMTAQDLDWNEPALIDQYHEGLSDHITQEEELSHLEV 240  
 QY 300 AKSLGALIGQCIHIERRLARAAAAARKPRSPRALVLPVPHIASHHQVDPTPEVGGARMRLTQ 359  
 D 300 AKSLGALIGQCIHIERRLARAAAAARKPRSPRALVLPVPHIASHHQVDPTPEVGGARMRLTQ 359  
 Db 241 AKSLGALIGQCIHIERRLARAAAAARKPRSPRALVLPVPHIASHHQVDPTPEVGGARMRLTQ 240  
 QY 360 EEKERRKLNLCYCGTGGHYADNCPAKAKSSPAGNSPAPL 401  
 D 360 EEKERRKLNLCYCGTGGHYADNCPAKAKSSPAGNSPAPL 401  
 Db 301 EEKERRKLNLCYCGTGGHYADNCPAKAKSSPAGNSPAPL 342

## RESULT 2

Q96A68 ID Q96A68 PRELIMINARY; PRT; 325 AA.  
 AC Q96A68;  
 DT 01-DEC-2001 (TrEMBLrel. 19, Created)  
 DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)  
 DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)  
 DE MEF3 LIKE 1 (PATERNALLY EXPRESSED GENE 10 ORF1).  
 GN MEF3L1 OR PEG10.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=TESTIS;  
 RA Nakamura Y., Furukawa Y.;  
 RL Submitted (SEP-2000) to the EMBL/GenBank/DBJ databases.  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=TESTIS;  
 RA Satoh S., Furukawa Y.;  
 RL "Isolation of MEF3 like gene 1.";  
 RT Submitted (SEP-2001) to the EMBL/GenBank/DBJ databases.  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=21218929; PubMed=11318613;  
 RA Ono R., Kobayashi S., Wagatsuma H., Aisaka K., Kohda T.,  
 RA Kaneko-Ishino T., Ishino F.;  
 RT "A Retrotransposon-Derived Gene, PEG10, Is a Novel Imprinted Gene  
 RT Located on Human Chromosome 7q21.";  
 RL Genomics 73:232-237(2001).  
 DR EMBL; AB049150; BAB68387.1; -;  
 DR EMBL; AB049834; BAB43951.1; -;  
 SQ SEQUENCE 325 AA; 36965 MW; 118E4CFAF97F2A76 CRC64;

Query Match 81.0%; Score 325; DB 4; Length 325;  
 Best Local Similarity 100.0%; Pred. No. 0;  
 Matches 325; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 77 MTERRRDSEIENLNREKVMKQSEENNLQSOVOKLTEENTTLREOVETPEDEDDDDIE 136  
 D 77 MTERRRDSEIENLNREKVMKQSEENNLQSOVOKLTEENTTLREOVETPEDEDDDDIE 136  
 Db 1 MTERRRDSEIENLNREKVMKQSEENNLQSOVOKLTEENTTLREOVETPEDEDDDDIE 60  
 QY 137 LRGAAAAAAPPPIIEECPCEDLPKFDGNDPDLAPFMAQCOIFMEKSTRDFSVDRVRVCF 196  
 D 137 LRGAAAAAAPPPIIEECPCEDLPKFDGNDPDLAPFMAQCOIFMEKSTRDFSVDRVRVCF 196  
 Db 61 LRGAAAAAAPPPIIEECPCEDLPKFDGNDPDLAPFMAQCOIFMEKSTRDFSVDRVRVCF 120  
 QY 197 VTSMTGTGAARWASAKLERSHYLMHNYPAFMEMKMHVFEDPQRREVAKRKIRRLROGMG 256  
 D 197 VTSMTGTGAARWASAKLERSHYLMHNYPAFMEMKMHVFEDPQRREVAKRKIRRLROGMG 256

Db 121 VTSMTGTGAARWASAKLERSHYLMHNYPAFMEMKMHVFEDPQRREVAKRKIRRLROGMG 180  
 QY 257 VIDYSNPFQMTAQDLDWNEPALIDQYHEGLSDHITQEEELSHLEVAKSLGALIGQCIHIERR 316  
 D 257 VIDYSNPFQMTAQDLDWNEPALIDQYHEGLSDHITQEEELSHLEVAKSLGALIGQCIHIERR 316  
 Db 181 VIDYSNPFQMTAQDLDWNEPALIDQYHEGLSDHITQEEELSHLEVAKSLGALIGQCIHIERR 240  
 QY 317 LARAAAAARKPRSPRALVLPVPHIASHHQVDPTPEVGGARMRLTOBEKERRKLNLCYCGT 376  
 D 317 LARAAAAARKPRSPRALVLPVPHIASHHQVDPTPEVGGARMRLTOBEKERRKLNLCYCGT 376  
 Db 241 LARAAAAARKPRSPRALVLPVPHIASHHQVDPTPEVGGARMRLTOBEKERRKLNLCYCGT 300  
 QY 377 GGHYADNCPAKAKSSPAGNSPAPL 401  
 D 377 GGHYADNCPAKAKSSPAGNSPAPL 401  
 Db 301 GGHYADNCPAKAKSSPAGNSPAPL 325

## RESULT 3

Q9EQ11 ID Q9EQ11 PRELIMINARY; PRT; 231 AA.  
 AC Q9EQ11;  
 DT 01-MAR-2001 (TrEMBLrel. 16, Created)  
 DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)  
 DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)  
 DE MYELIN EXPRESSION FACTOR-3-LIKE PROTEIN (FRAGMENT).  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OX NCBI\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=C57BL/6;  
 RX Wolff J.-N., Koerting C., Scharltl M.;  
 RL "Ty3/Gypsy retrotransposon fossils in mammalian genomes: did they  
 RL evolve into new cellular functions?";  
 DR Mol. Biol. Evol. 18:286-270(2001).  
 DR EMBL; AF302691; AAG39979.1; -;  
 DR InterPro; IPR001878; Znf\_CCHC.  
 DR Pfam; PF00098; zf-CCHC; 1.  
 DR SMART; SM00343; Znf\_C2HC; 1.  
 KW Zinc-finger.  
 FT NON\_TER 1 1  
 SQ SEQUENCE 231 AA; 26171 MW; DEAB2A2B624F3974 CRC64;

Query Match 3.7%; Score 15; DB 11; Length 231;  
 Best Local Similarity 100.0%; Pred. No. 3e-06;  
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 259 DYSNPFQMTAQDLDW 273  
 D 259 DYSNPFQMTAQDLDW 107  
 Db 93 DYSNPFQMTAQDLDW 107

## RESULT 4

Q9AYL5 ID Q9AYL5 PRELIMINARY; PRT; 212 AA.  
 AC Q9AYL5;  
 DT 01-JUN-2001 (TrEMBLrel. 17, Created)  
 DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)  
 DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)  
 DE HYPOTHETICAL 23.1 KDA PROTEIN.  
 GN OSJNBA0003019.2.  
 OS Oryza sativa (Rice).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;  
 OC Ehrhartoideae; Oryzaceae; Oryza.  
 OX NCBI\_TaxID=4530;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=CV, NIPPONBARE;  
 RA Buell C.R., Yuan Q., Moffat K.S., Hill J.N., Burr P.C., Hsiao J.,  
 RA Zismann V., Pal G., Bowman C.L., Fujii C.Y., Vanaken S.E.,  
 RA Bowman C.L., Craven B., Utterback T.R., Khalak H., Feldblum T.V.,



RA Quackenbush J., White O., Salzberg S.L., Fraser C.M.;  
 RT "Oryza sativa chromosome 10 BAC OSJNBa0003019 genomic sequence.";  
 RL Submitted (FEB-2001) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: AC060755; AAK00445.1; -;  
 DR InterPro: IPR000916; Bel\_v\_1.  
 DR ProDom: PD000531; Bel\_v\_1; 1.  
 KW Hypothetical protein.  
 SQ SEQUENCE 212 AA; 23084 MW; BBA1D03553465083 CRC64;

Query Match 2.5%; Score 10; DB 10; Length 212;  
 Best Local Similarity 100.0%; Pred. No. 0.21;  
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 140 AAAAAA PPPP 149  
 DB 197 AAAAAA PPPP 206  
 |||||

RESULT 5  
 Q9XFBL PRELIMINARY; PRT; 240 AA.  
 AC Q9XFBL  
 DT 01-NOV-1999 (TREMBlrel. 12, Created)  
 DT 01-NOV-1999 (TREMBlrel. 12, Last sequence update)  
 DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)  
 DE YABBY3 (At4G00180/FGN15\_22).  
 GN YABBY3  
 OS Arabidopsis thaliana (Mouse-ear cress).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;  
 OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.  
 OX NCBI\_TaxID=3702;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=99387979; PubMed=10457020;  
 RA Siegfried K.R., Eshed Y., Baum S.F., Otsuga D., Drews G.N.,  
 Bowman J.L.;  
 RT "Members of the YABBY gene family specify abaxial cell fate in  
 RT Arabidopsis.";  
 RL Development 126:4117-4128(1999).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RA Cheuk R., Chen H., Kim C.J., Koesema E., Meyers M.C., Shinn P.,  
 Tracy S.E., Banh J., Bowser L., Carinci P., Chung M.K.,  
 Goldsmith A.D., Hayashizaki Y., Ishida J., Jones T., Kamiya A.,  
 Karlin-Neumann G., Kawai J., Lam B., Lee J.M., Lin J., Liu S.X.,  
 Miranda M., Narusaka M., Nguyen M., Palm C.J., Pham P.K., Quach H.L.,  
 Sakano H., Sakurai T., Satou M., Seki M., Southwick A., Tang C.C.,  
 Toriumi M., Yamada K., Yu G., Shinozaki K., Davis R.W., Theologis A.,  
 Ecker J.R.;  
 RT "Arabidopsis cDNA clones";  
 RL Submitted (MAY-2001) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: AF136540; AAD33717.1; -;  
 DR EMBL: AY037186; AAK59771.1; -;  
 SQ SEQUENCE 240 AA; 26338 MW; 4D90DA549B54A9AC CRC64;

Query Match 2.5%; Score 10; DB 10; Length 240;  
 Best Local Similarity 100.0%; Pred. No. 0.23;  
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 48 PPPPPPPPPN 57  
 DB 84 PPPPPPPPPN 93  
 |||||

RESULT 6  
 Q20327 PRELIMINARY; PRT; 250 AA.  
 ID Q20327  
 AC Q20327  
 DT 01-NOV-1996 (TREMBlrel. 01, Created)  
 DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)

DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)  
 DE HYPOTHETICAL 27.7 KDA PROTEIN.  
 GN F42C5.7.  
 OS Caenorhabditis elegans.  
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;  
 OC Rhabditidae; Peloderinae; Caenorhabditis.  
 OX NCBI\_TaxID=6239;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-BRISTOL N2;  
 RX MEDLINE=99069613; PubMed=9851916;  
 RA None;  
 RT "Genome sequence of the nematode C. elegans: a platform for  
 RT investigating biology. The C. elegans Sequencing Consortium.";  
 RL Science 282:2012-2018(1998).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-BRISTOL N2;  
 RA Du Z.;  
 RT "The sequence of C. elegans cosmid F42C5.";  
 RL Submitted (NOV-1995) to the EMBL/GenBank/DBJ databases.  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-BRISTOL N2;  
 RA Waterston R.;  
 RT "Direct Submission.";  
 RL Submitted (JUN-2001) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: U40799; AAA81484.1; -;  
 DR InterPro: IPR001467; Molybdopterin.  
 DR PROSITE: PS00551; MOLYBDOPTERIN\_PROK\_1; UNKNOWN\_1.  
 KW Hypothetical protein.  
 SQ SEQUENCE 250 AA; 27685 MW; 477A9EC12E3C3585 CRC64;

Query Match 2.5%; Score 10; DB 5; Length 250;  
 Best Local Similarity 100.0%; Pred. No. 0.24;  
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 47 CPPPPPPPPP 56  
 DB 102 CPPPPPPPPP 111  
 |||||

RESULT 7  
 Q9DGN3 PRELIMINARY; PRT; 268 AA.  
 ID Q9DGN3  
 AC Q9DGN3  
 DT 01-MAR-2001 (TREMBlrel. 16, Created)  
 DT 01-MAR-2001 (TREMBlrel. 16, Last sequence update)  
 DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)  
 DE ESPIN (FRAGMENT).  
 OS Gallus gallus (Chicken).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;  
 OC Gallus.  
 OX NCBI\_TaxID=9031;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=COCHLEA;  
 RX MEDLINE=20427267; PubMed=10975527;  
 RA Zheng L., Sekerkova G., Vranich K., Tilney L.G., Mugnaini E.,  
 Bartles J.R.;  
 RT "The Deaf Jerker Mouse Has a Mutation in the Gene Encoding the Espin  
 RT Actin-Bundling Proteins of Hair Cell Stereocilia and Lacks Espins.";  
 RL Cell 102:377-385(2000).  
 DR EMBL: AF239885; AAF98133.1; -;  
 DR InterPro: IPR000381; Inhibin\_beta.  
 DR InterPro: IPR002965; P\_rich\_extensn.  
 DR InterPro: IPR000633; Vinculin\_2.  
 DR InterPro: IPR003124; WH2.  
 DR Pfam: PF02205; WH2; 1.  
 DR PRINTS: PR00671; INHIBINBB.  
 DR PRINTS: PR01217; PRICHEXTNSN.

```

DR PRINTS: PR00806; VINCU LIN.
DR SMART; SM00246; WH2; 1.
FT NON_TER 1
SQ SEQUENCE 268 AA; 29103 MW; F815B4828B2BB9A2 CRC64;

Query Match 2.5%; Score 10; DB 13; Length 268;
Best Local Similarity 100.0%; Pred. No. 0.25;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 47 C P P P P P P P P P P 56
   | | | | | | | | | |
Db 26 C P P P P P P P P P 35

RESULT 8
Q9SNS1 ID Q9SNS1 PRELIMINARY; PRT; 276 AA.
AC Q9SNS1;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE HYPOTHETICAL PROTEIN.
OS Oryza sativa (Rice).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC Ehrhartoideae; Oryzeae; Oryza.
OX NCBI_TaxID=4530;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CV. NIPPONBARE;
RA Sasaki T., Matsumoto T., Yamamoto K.;
RT "Oryza sativa nipponbare(GA3) genomic DNA, chromosome 6, PAC
   clone:P0535G04."
RL Submitted (AUG-1999) to the EMBL/GenBank/DBJ databases.
DE EMBL; AP000399; BA83555.1; -.
KW Hypothetical protein.
SQ SEQUENCE 276 AA; 29463 MW; 62B4AA1069536447 CRC64;

Query Match 2.5%; Score 10; DB 10; Length 276;
Best Local Similarity 100.0%; Pred. No. 0.26;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 140 A A A A A A P P P P 149
   | | | | | | | | | |
Db 79 A A A A A A P P P P 88

RESULT 9
Q9CUF6 ID Q9CUF6 PRELIMINARY; PRT; 384 AA.
AC Q9CUF6;
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE 4930577N17RIK PROTEIN (FRAGMENT).
GN 4930577N17RIK.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=TESTIS;
RA MEDLINE=21085660; PubMed=11217851;
RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I.,
RA Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
RA Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,

Schröml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,
Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
Nordone P., Ring B., Schoenwald M., Rodriguez I., Sakamoto N.,
Sasaki H., Sato K., Schenbach C., Seya T., Shibata Y., Storch K.-F.,
Suzuki H., Toyooka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,
Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohetsuki S.,
Hayashizaki Y.;
FT *Functional annotation of a full-length mouse cDNA collection.*;
RL Nature 409:685-690(2001).
DE EMBL; AK016297; BAB30183.1; -.
DR MGD; MGI:1914996; 4930577N17RIK.
FT NON_TER 1
SQ SEQUENCE 384 AA; 41209 MW; E9856EA599F28262 CRC64;

Query Match 2.5%; Score 10; DB 11; Length 384;
Best Local Similarity 100.0%; Pred. No. 0.34;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 140 A A A A A A P P P P 149
   | | | | | | | | | |
Db 62 A A A A A A P P P P 71

RESULT 10
Q9SBM1 ID Q9SBM1 PRELIMINARY; PRT; 409 AA.
AC Q9SBM1;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE HYDROXYPROLINE-RICH GLYCOPROTEIN DZ-HRGP PRECURSOR.
GN HRGP GENE.
OS Volvox carteri f. nagariensis.
OC Eukaryota; Viridiplantae; Chlorophyta; Chlorophyceae; Volvocales;
OC Volvocaceae; Volvox.
OX NCBI_TaxID=3068;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=HK10;
RX MEDLINE=20044763; PubMed=10574980;
RA Ender F., Hallmann A., Amon P., Sumper M.;
RT "Response to the sexual pheromone and wounding in the green alga
   Volvox: induction of an extracellular glycoprotein consisting almost
   exclusively of hydroxyproline.";
RL J. Biol. Chem. 274:35023-35028(1999).
DR EMBL; AJ242540; CAB62280.1; -.
DR InterPro; IPR003882; Pistil_extensin.
DR InterPro; IPR002965; P_rich_extensin.
DR PRINTS; PR01217; PRICHEXTENSIN.
DR PRINTS; PR01218; PSTLEXTENSIN.
KW Signal.
FT SIGNAL 1
FT CHAIN 18 409 HYDROXYPROLINE-RICH GLYCOPROTEIN DZ-HRGP.
SQ SEQUENCE 409 AA; 41547 MW; CD0749C6AF02BD74 CRC64;

Query Match 2.5%; Score 10; DB 10; Length 409;
Best Local Similarity 100.0%; Pred. No. 0.36;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 48 P P P P P P P P P P 57
   | | | | | | | | | |
Db 129 P P P P P P P P P P 138

RESULT 11
Q9FXA1 ID Q9FXA1 PRELIMINARY; PRT; 494 AA.

```

AC Q9FXA1;  
 DT 01-MAR-2001 (TREMBLrel. 16, Created)  
 DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)  
 DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)  
 DE F14J22.4 PROTEIN.  
 GN F14J22.4.  
 OS Arabidopsis thaliana (Mouse-ear cress).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;  
 OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.  
 OX NCBI\_TaxID=3702;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Federspiel N.A., Palm C.J., Conway A.B., Conn L., Hansen N.F.,  
 RA Alfati H., Araujo R., Huizar L., Rowley D., Buehler E., Dunn P.,  
 RA Gonzalez A., Kremenetskaia I., Kim C., Lenz C., Li J., Liu S.,  
 RA Luros S., Schwartz J., Shinn P., Toriumi M., Vysotskaia V.S.,  
 RA Walker M., Yu G., Ecker J., Theologis A., Davis R.W.;  
 RL Submitted (SEP-2000) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AC011807; AAG13059.1; -;  
 DR InterPro: IPR001611; LRR.  
 DR InterPro: IPR003592; LRR.out.  
 DR InterPro: IPR002965; P-rich\_extensn.  
 DR Pfam: PF00360; LRR; 8.  
 DR PRINTS; PRO0019; LEURICHRPT.  
 DR PRINTS; PRO1217; PRICHEXTENS.  
 DR SMART; SM00370; LRR; 5.  
 SQ SEQUENCE 494 AA; 54407 MW; BDC0011EEFD5C2E4 CRC64;  
 Query Match 2.5%; Score 10; DB 10; Length 494;  
 Best Local Similarity 100.0%; Pred. No. 0.42;  
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 46 DCPPPPPPPPP 55  
 DB 60 DCPPPPPPPPP 69  
 RESULT 12  
 O61701 PRELIMINARY; PRT; 1018 AA.  
 AC O61701;  
 DT 01-AUG-1998 (TREMBLrel. 07, Created)  
 DT 01-AUG-1998 (TREMBLrel. 07, Last sequence update)  
 DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)  
 DE HYPOTHETICAL 112.9 KDA PROTEIN (FRAGMENT).  
 GN CYK-1.  
 OS Caenorhabditis elegans.  
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;  
 OC Rhabditidae; Peloderinae; Caenorhabditis.  
 OX NCBI\_TaxID=6239;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Swann K.A., Severson A.F., Carter J.C., Martin P.R., Schnabel H.,  
 RA Schnabel R., Bowerman B.;  
 RT "cyk-1: a C. elegans FH gene required for a late step in embryonic  
 RT cytokinesis.";  
 RL J. Cell Sci. 0:0-0(1998).  
 DR EMBL; AF062008; AAC17501.1; -;  
 DR InterPro: IPR003104; FH2.  
 DR InterPro: IPR000381; Inhibin\_betaB.  
 DR Pfam: PF02181; FH2; 1.  
 DR PRINTS; PR00671; INHIBINB.  
 DR SMART; SM00498; FH2; 1.  
 KW Hypothetical protein.  
 FT NON\_TER 1  
 SQ SEQUENCE 1018 AA; 112944 MW; 07DA53F14B000189 CRC64;  
 Query Match 2.5%; Score 10; DB 5; Length 1018;  
 Best Local Similarity 100.0%; Pred. No. 0.77;

Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 47 CPPPPPPPPPP 56  
 DB 351 CPPPPPPPPPP 360  
 RESULT 13  
 Q95ZV4 PRELIMINARY; PRT; 1432 AA.  
 AC Q95ZV4;  
 DT 01-DEC-2001 (TREMBLrel. 19, Created)  
 DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)  
 DE 01-DEC-2001 (TREMBLrel. 19, Last annotation update)  
 DE HYPOTHETICAL PROTEIN F11H8.4.  
 GN F11H8.4.  
 OS Caenorhabditis elegans.  
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;  
 OC Rhabditidae; Peloderinae; Caenorhabditis.  
 OX NCBI\_TaxID=6239;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-BRISTOL N2;  
 RA MEDLINE=99069613; PubMed=9851916;  
 RX None;  
 RT "Genome sequence of the nematode C. elegans: a platform for  
 RT investigating biology. The C. elegans Sequencing Consortium.";  
 RL Science 282:2012-2018(1998).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-BRISTOL N2;  
 RA Menezes S., Menezes S.;  
 RT "The sequence of C. elegans cosmid F11H8.";  
 RL Submitted (NOV-1995) to the EMBL/GenBank/DBJ databases.  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-BRISTOL N2;  
 RA Waterston R.;  
 RT "Direct Submission.";  
 RL Submitted (JUN-2001) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; U40187; AAA81161.2; -;  
 SQ SEQUENCE 1432 AA; 159440 MW; 5560B3C205DBCD37 CRC64;  
 Query Match 2.5%; Score 10; DB 5; Length 1432;  
 Best Local Similarity 100.0%; Pred. No. 1;  
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 47 CPPPPPPPPPP 56  
 DB 768 CPPPPPPPPPP 777  
 RESULT 14  
 O02832 PRELIMINARY; PRT; 33 AA.  
 AC O02832;  
 DT 01-JUL-1997 (TREMBLrel. 04, Created)  
 DT 01-JUL-1997 (TREMBLrel. 04, Last sequence update)  
 DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)  
 DE HUNTINGTIN PROTEIN (FRAGMENT).  
 GN IT15.  
 OS Gorilla gorilla (gorilla).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Gorilla.  
 OX NCBI\_TaxID=9593;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=96326790; PubMed=8766138;  
 RA Pecheux C., Gall A.L., Kaplan J.C., Dode C.;  
 RT "Sequence analysis of the CAG triplet repeats region in the Huntington  
 RT disease gene (IT15) in several mammalian species.";  
 RL Ann. Genet. 39:81-86(1996).

DR EMBL; S83377; AAB50771.1; -.

FT NON\_TER 1

SQ SEQUENCE 33 AA; 3393 MW; 44F558CE5636722A CRC64;

Query Match 2.2%; Score 9; DB 6; Length 33;  
Best Local Similarity 100.0%; Pred. No. 0.42;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 48 PPPPPPPPP 56

|||||

DB 20 PPPPPPPPP 28

RESULT 15

Q9VWS6

ID Q9VWS6 PRELIMINARY; PRT; 48 AA.

AC Q9VWS6;

DT 01-MAY-2000 (TREMBlrel. 13, Created)

DT 01-MAR-2001 (TREMBlrel. 16, Last sequence update)

DT 01-MAR-2001 (TREMBlrel. 16, Last annotation update)

DE CG15041 PROTEIN.

GN CG15041

OS Drosophila melanogaster (Fruit fly).

OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;

OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;

OC Ephydroidea; Drosophilidae; Drosophila.

OX NCBI\_TaxID=7227;

RN {}

RP SEQUENCE FROM N.A.

RC STRAIN=BERKELEY;

RX MEDLINE=20196006; PubMed=10731132;

RA Adams M.D., Celnikier S.E., Holt R.A., Evans C.A., Gocayne J.D.,

RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,

RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,

RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,

RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,

RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,

RA Aril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,

RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,

RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,

RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brotter P.,

RA Burlis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,

RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,

RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,

RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,

RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,

RA Foster C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,

RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,

RA Hostin D., Houston K.A., Howland T.J., Hernandez J.R., Houck J.,

RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,

RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,

RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,

RA Liu X., Mattel B., McIntosh T.C., McLeod M.P., McPherson D.,

RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,

RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,

RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,

RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,

RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,

RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,

RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,

RA Svirkas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,

RA Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissbach J.,

RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,

RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,

RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,

RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;

RT "The genome sequence of Drosophila melanogaster.";

RL Science 287:2185-2195(2000).

DR EMBL; AE003509; AAF48862.2; -.

DR FlyBase; FBgn0030939; CG15041

SQ SEQUENCE 48 AA; 4976 MW; 17291770CAF8F177 CRC64;

Query Match 2.2%; Score 9; DB 5; Length 48;  
Best Local Similarity 100.0%; Pred. No. 0.57;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 48 PPPPPPPPP 56

|||||

DB 25 PPPPPPPPP 33

Search completed: August 13, 2002, 15:15:05  
Job time: 141 sec



**THIS PAGE BLANK (USPTO)**